

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:14:52 ; Search time 73 Seconds
(without alignments)
848.788 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKKYQYFALALCALALAG.....ERAGLRHGDHLAVRASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	A.GeneSeq.101002.*
1:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23:	/SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2353	100.0	465	21	NGSP polypeptide o
2	2317	98.5	499	21	Neisseria gonorrhoe
3	2254	95.8	499	21	Neisseria meningit
4	2248	95.5	499	21	Neisseria meningit
5	2244	95.4	499	21	Neisseria meningit
6	2234.5	95.0	498	21	Neisseria meningit
7	2202	93.6	499	21	Neisseria meningit
8	2197	93.4	499	21	Neisseria meningit
9	2130.5	90.5	475	21	Neisseria meningit
10	2092	88.9	414	21	NGSP polypeptide o

11	2002	85.1	448	21	AAV70409
12	1832	77.9	370	20	AAV52996
13	962.5	40.9	460	20	AAV29294
14	731.5	31.1	503	17	AAV7434
15	705	30.0	474	22	AAV78605
16	685	29.1	491	20	AAV37309
17	668.5	28.4	488	13	ABR90581
18	633.5	26.9	475	12	AAV4625
19	632	26.9	463	17	AAV2690
20	629	26.7	437	21	AAV1834
21	627.5	26.7	455	22	AAV9896
22	624	26.5	437	21	AAV11837
23	624	26.5	437	21	AAV11838
24	623	26.4	437	21	AAV11836
25	622	26.4	437	21	AAV11835
26	603.5	25.6	491	19	AAV56771
27	573.5	24.4	443	20	AAV9850
28	565.5	24.0	451	20	AAV89977
29	559.5	23.8	387	19	AAV98445
30	542.5	23.1	433	21	AAV59353
31	542.5	23.1	433	21	AAV59354
32	533	22.7	433	21	AAV59354
33	453	19.3	397	21	AAV81772
34	449.5	19.1	320	21	AAV80962
35	448	19.0	382	21	AAV80958
36	448	19.0	394	21	AAV5193
37	448	19.0	394	21	AAV5193
38	448	19.0	451	21	AAV80957
39	448	19.0	480	18	AAV2849
40	448	19.0	480	21	AAV80956
41	448	19.0	480	21	AAV15191
42	448	19.0	517	23	ABP41178
43	447.5	19.0	355	22	AAV98947
44	445	18.9	480	21	AAV80964
45	443.5	18.8	405	21	AAV58844

ALIGNMENTS

RESULT 1	AAV83150	standard; Protein; 465 AA.
ID	AAV83150	
AC	AAV83150;	
XX		
DT	24-JUL-2000	(first entry)
DE	NGSP polypeptide of Neisseria gonorrhoeae.	
XX		
KW	NGSP: polypeptide; peptide; vaccine; immune response; antibody;	
KW	cellular matrix; adherence domain; ligand; detection; diagnosis;	
KW	screening; probe; primer; prophylaxis; therapy.	
XX		
OS	Neisseria gonorrhoeae.	
XX		
PN	WQ200012133-A1.	
XX		
PD	09-MAR-2000.	
XX		
PF	01-SEP-1999;	99WO-US20070.
XX		
PR	01-SEP-1998;	98US-0098685.
XX		
PA	(ANTE-) ANTEX BIOLOGICS INC.	
XX		
PI	Jackson WJ, Harris AM;	
XX		
DR	WPI; 2000-237782/20.	
XX		
PT	N-PSDB; AA293414.	
XX		
PT	Non-cytosolic NGSP polypeptide and polynucleotide sequence from	
PT	Neisseria useful for diagnosis, prevention or treatment of Neisseria	

Neisseria meningit
Neisseria meningit
Protein encoded by
Heat shock protein
Lawsonia intracell
Protein involved i
Chlamydia pneumon
S. typhimurium Htr
H. influenzae SB33
Haemophilus influe
E. coli growth and
Haemophilus influe
Haemophilus influe
Haemophilus influe
Bacterioides coli s
Antigen 1 from clu
Protein encoded by
H. pylori GHPO 536
BASB01 protein se
BASB01 protein se
Mycobacterium tube
Streptococcus pneu
Human HtrA (high t
Human HtrA (high t
C-terminal domain
Human HtrA (high t
Osteoblast like ce
Human HtrA (high t
Human serine prote
Human ovarian anti
E. coli growth and
Human HtrA (high t
Breast and ovarian

PT infections
 XX Claim 5; Page 61-62; 68pp; English.
 PS
 XX The NGSP polypeptide of *N. gonorrhoeae* has conserved Arg-Gly-Asp and
 CC Arg-Gly-Asn motifs near the C-terminus which function as adherence
 CC domains for extracellular matrix proteins. Using the NGSP polypeptide
 CC as a vaccine produces antibodies which inhibit binding of *N.*
 CC gonorrhoeae to the host's cellular matrix reducing attachment and/or
 CC subsequent invasion. The NGSP polypeptide and its peptide fragments
 CC can be used to immunise an animal and produce an immune response.
 CC They can also be used as ligands to detect antibodies elicited in
 CC response to *Neisseria* infections and also as antigens or immunogens
 CC for inducing *Neisseria*-specific antibodies which are useful in
 CC immunoassays to detect *Neisseria* in biological specimens. Nucleotides
 CC encoding NGSP or its fragments can be used as probes to identify
 CC *Neisseria* in biological specimens by hybridization or polymerase
 CC chain reaction amplification. The NGSP polypeptide can also be used
 CC in screening assays to identify agents and compounds which useful as
 CC diagnostic, prophylactic or therapeutic agents against *Neisseria*
 CC infection.

Sequence 465 AA;
 SQ
 Query Match 100.0%; Score 2353; DB 21; Length 465;
 Best Local Similarity 100.0%; Pred. No. 1,6e-190;
 Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VFKKYQYFALAAACALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 1 VFKKYQYFALAAACALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 1 VFKKYQYFALAAACALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 121 GGLNFGSGFIISKNGYLLTNTHVAVGMSIKVLNDKREYAKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYLLTNTHVAVGMSIKVLNDKREYAKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYLLTNTHVAVGMSIKVLNDKREYAKLIGSDVQSDVALLKIDA 180
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGDNSTAGIVSAKGRSLPNESYTPFIOTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGDNSTAGIVSAKGRSLPNESYTPFIOTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGDNSTAGIVSAKGRSLPNESYTPFIOTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGDNSTAGIVSAKGRSLPNESYTPFIOTDVA 240
 241 INPNSGGPLFNLKQGVVGINSGIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 241 INPNSGGPLFNLKQGVVGINSGIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 241 INPNSGGPLFNLKQGVVGINSGIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 301 LGVLIQEVSYGLAQSGFLDASGALIAKILPGSPAERAGIQAGDIVSLDGEIRSSGDL 360
 301 LGVLIQEVSYGLAQSGFLDASGALIAKILPGSPAERAGIQAGDIVSLDGEIRSSGDL 360
 301 LGVLIQEVSYGLAQSGFLDASGALIAKILPGSPAERAGIQAGDIVSLDGEIRSSGDL 360
 361 PVMVGAITPEKEVSLGVWRKGEETITAKIGNAAEHTGASKTDBAPTEQSGTFVYES 420
 361 PVMVGAITPEKEVSLGVWRKGEETITAKIGNAAEHTGASKTDBAPTEQSGTFVYES 420
 361 PVMVGAITPEKEVSLGVWRKGEETITAKIGNAAEHTGASKTDBAPTEQSGTFVYES 420
 421 AGTTLQTHDSSGKHLVVRVSDAERAGLRHGEILLAVRASPQ 465
 421 AGTTLQTHDSSGKHLVVRVSDAERAGLRHGEILLAVRASPQ 465
 421 AGTTLQTHDSSGKHLVVRVSDAERAGLRHGEILLAVRASPQ 465

RESULT 2
 ID AAY75748
 ID AAY75748 standard; Protein: 499 AA.
 AC AAY75748;
 DT 21-MAR-2000 (first entry)
 DE *Neisseria gonorrhoeae* ORF 986 protein sequence SEQ ID NO:2968.
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnostic; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy.
 XX
 OS *Neisseria gonorrhoeae*.
 XX
 XX WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-P8DB; AA254510.
 XX
 PT Novel *Neisseria* polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 1390; 1453pp; English.
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccine, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX

Sequence 499 AA;
 SQ
 Query Match 98.5%; Score 2317; DB 21; Length 499;
 Best Local Similarity 99.1%; Pred. No. 2e-187;
 Matches 459; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 VFKKYQYFALAAACALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 1 VFKKYQYFALAAACALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 1 VFKKYQYFALAAACALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 121 GGLNFGSGFIISKNGYLLTNTHVAVGMSIKVLNDKREYAKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYLLTNTHVAVGMSIKVLNDKREYAKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYLLTNTHVAVGMSIKVLNDKREYAKLIGSDVQSDVALLKIDA 180
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGDNSTAGIVSAKGRSLPNESYTPFIOTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGDNSTAGIVSAKGRSLPNESYTPFIOTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGDNSTAGIVSAKGRSLPNESYTPFIOTDVA 240
 241 INPNSGGPLFNLKQGVVGINSGIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 241 INPNSGGPLFNLKQGVVGINSGIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 241 INPNSGGPLFNLKQGVVGINSGIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300

QY 301 LGVITIEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLSLDGSEIRSSGDL 360
 Db 301 LGVITIEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLSLDGSEIRSSGDL 360
 QY 361 PVMVGAITPGKEVSLGVWRKGEITIKAKLGNAAEHTGASKTDEAPYTEQSGTFSVES 420
 Db 361 PVMVGAITPGKEVSLGVWRKGEITIKAKLGNAAEHTGASKTDEAPYTEQSGTFSVES 420
 QY 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVRASP 463
 Db 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVGQVP 463

RESULT 3
 ID AAY52995 standard; Protein: 499 AA.
 AAY52995;

21-FEB-2000 (first entry)

Neisseria meningitidis strain H44/76 BASB013 protein sequence.
 DE Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
 XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
 KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
 KM invasive bacterial disease; antibacterial.
 XX

Neisseria meningitidis.

W09955872-A1.

04-NOV-1999.

20-APR-1999; 99WO-EP02765.

23-APR-1998; 98GB-0008734.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI; 2000-052809/04.

N-PSDB; AAZ33307.

Novel polynucleotides and polypeptides from Neisseria meningitis used to prepare vaccines against bacterial infections

Claim 3; Page 79-81; 94pp; English.

CC The present sequence represents a BASB013 polypeptide isolated from
 CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be
 CC employed as research reagents and material for the discovery of
 CC treatments and diagnostics for diseases, particularly human diseases.
 CC They can be used for diagnosis of disease, staging of disease, or
 CC determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and
 CC for screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC agonists and antagonists. The polypeptides, antibodies, agonists and
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases such as upper respiratory tract infection,
 CC invasive bacterial diseases such as bacteraemia and meningitis, and for
 CC the development and screening of antibacterial drugs. They are also used
 CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins
 CC on in-dwelling devices, or to extracellular proteins on wounds, and to
 CC thus prevent tissue damage and/or block the normal progression of
 CC pathogens in infections initiated other than by the implantation of
 CC in-dwelling devices or by other surgical techniques.

SEQ Sequence 499 AA;

Query Match 95.8%; Score 2254; DB 21; Length 499;
 Best Local Similarity 96.1%; Pred. No. 4.4e-182;
 Matches 445; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 VFKKYQYFALALCAALLAGCEKASFFGADKKEASFVERIEHTKDDGSVSMLLPDEFQOL 60
 Db 1 MFKKYQYFALALCAALLAGCEKASFFGADKKEASFVERIEHTKDDGSVSMLLPDEFQOL 60
 QY 61 VQSEGPVAVNIQAAPAPRTQNGSGNAENDSDPIANDPFEFFKLVNMPETIPOEADD 120
 Db 61 VQSEGPVAVNIQAAPAPRTQNGSGNAENDSDPIANDPFEFFKLVNMPETIPOEADD 120
 QY 121 GGLNFGSGFIISKNGIYILTNTHVAGMGIKYLNDKREYTKLIGSDVQSDVALLIKIDA 180
 Db 121 GGLNFGSGFIISKNGIYILTNTHVAGMGIKYLNDKREYTKLIGSDVQSDVALLIKIDA 180
 QY 181 TEELPVVKIGNPKNLKPGEWVAIAGPFGFDSVTAAGIVSAKGRSLPNESYTPFIQTDVA 240
 Db 181 TEELPVVKIGNPKNLKPGEWVAIAGPFGFDSVTAAGIVSAKGRSLPNESYTPFIQTDVA 240
 QY 241 INPNSGGFELFKGQVGINSGIYSRSGFMGISFAIPIDVAMNVAEQLKNTGKVRQ 300
 Db 241 INPNSGGFELFKGQVGINSGIYSRSGFMGISFAIPIDVAMNVAEQLKNTGKVRQ 300
 QY 301 LGVITIEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLSLDGSEIRSSGDL 360
 Db 301 LGVITIEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLSLDGSEIRSSGDL 360
 QY 361 PVMVGAITPGKEVSLGVWRKGEITIKAKLGNAAEHTGASKTDEAPYTEQSGTFSVES 420
 Db 361 PVMVGAITPGKEVSLGVWRKGEITIKAKLGNAAEHTGASKTDEAPYTEQSGTFSVES 420
 QY 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVRASP 463
 Db 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVGQVP 463

RESULT 4

ID AAY75749 standard; Protein: 499 AA.

AAY75749;

21-MAR-2000 (first entry)

Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2970.

Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.

Neisseria meningitidis.

W09957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.

31-JUL-1998; 98US-0094869.

02-SEP-1998; 98US-0098994.

09-OCT-1998; 98US-0099062.

09-OCT-1998; 98US-0103749.

09-OCT-1998; 98US-0103794.

25-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.

(GENO-) INST GENOMIC RES.


```

Qy 241 INPNSGGPLFNLKQVVGINSQIYSRSQGFMSFALPIDVAMNVABQLKNTGKYVRGQ 300
Db 241 INPNSGGPLFNLKQVVGINSQIYSRSQGFMSFALPIDVAMNVABQLKNTGKYVRGQ 300
Qy 301 LGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIYLSLDGGEIRSSGDL 360
Db 301 LGVIIQEVSYGLAOSFGLDKAGALIAKILPGSPAERAGLRAGDIYLSLDGGEIRSSGDL 360
Qy 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOGSTFSVES 420
Db 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOGSTFSVES 420
Qy 421 AGITLQHTDSSGKHLVVVRVSDPAERAGLRHGEDEILAVRASP 463
Db 421 AGITLQHTDSSGKHLVVVRVSDPAERAGLRHGEDEILAVRASP 463
Qy 421 AGITLQHTDSSGKHLVVVRVSDPAERAGLRHGEDEILAVRASP 463
Db 421 AGITLQHTDSSGKHLVVVRVSDPAERAGLRHGEDEILAVRASP 463

RESULT 6
0413
AA770413 standard; Protein; 498 AA.
AC AA770413;
XX 03-JUN-2000 (first entry)
XX
DE Neisseria meningitidis NMASP protein-2.
XX
XX NMASP, non-cytosolic; antibacterial; antiinflammatory; cytotoxic;
XX anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;
XX Neisserial infection; meningitidis; septicemia.
XX
XX OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT Misc-difference 8 /note= "Encoded by TTC"
FT Misc-difference 17 /note= "Encoded by TTG"
FT Misc-difference 22 /note= "Encoded by GAA"
FT Misc-difference 42 /note= "Encoded by GAA"
FT Misc-difference 58 /note= "Encoded by GAA"
FT Misc-difference 84 /note= "Encoded by GCC"
FT Misc-difference 134 /note= "Encoded by GGC"
FT Misc-difference 145 /note= "Encoded by AAC"
FT Misc-difference 194 /note= "Encoded by GCC"
FT Misc-difference 218 /note= "Encoded by AAT"
FT Misc-difference 321 /note= "Encoded by GGCATC"
FT Misc-difference 340 /note= "Encoded by AGC"
FT Misc-difference 387 /note= "Encoded by AGG"
FT Misc-difference 396 /note= "Encoded by GCC"
FT Misc-difference 433 /note= "Encoded by ACC"
FT Misc-difference 451 /note= "Encoded by AAA"
FT Misc-difference 459 /note= "Encoded by CAC"
FT Misc-difference 460 /note= "Encoded by AGG"
FT Misc-difference 461 /note= "Encoded by GCA"
FT /note= "Encoded by AGT"

```

```

FT Misc-difference 463 /note= "Encoded by CGT"
FT Misc-difference 464 /note= "Encoded by CAA"
FT Misc-difference 465.498 /note= "Encoded by GAGCAAGTCCCGTCGAA"
FT W0200012535-A2.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WC0-US19663.
XX
XX 01-SEP-1998; 98US-0098685.
XX
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Jackson WJ, Harris AM;
XX
XX WPI; 2000-256581/22.
XX
XX N-PSDB; AA251538.
XX
XX Neisseria meningitidis NMASP polypeptide, nucleotide sequences and
XX antibodies, useful in vaccines against infection -
XX
XX Claim 5; Page 70-71; 75pp; English.
XX
XX The present sequence is the Neisseria meningitidis NMASP protein.
XX NMASP is a non-cytosolic protein, with antibacterial and
XX antiinflammatory activity. It shows sequence similarity to E. coli
XX DsgP (HtrA) protein. NMASP proteins can be used as ligands to detect
XX antibodies elicited in response to N. meningitidis infections. Cytotoxic
XX anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA
XX may be used for diagnosis, therapy or prophylaxis of Neisserial
XX infections such as, bacterial meningitidis and septicemia.
XX
XX Sequence 498 AA;
XX
XX Query Match 95.0%; Score 2234.5; DB 21; Length 498;
XX Best Local Similarity 95.9%; Pred. No. 1.9e-180;
XX Matches 444; Conservative 5; Mismatches 13; Indels 1; Gaps 1;
Qy 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFEVERIEHTKDDGSVMTLPDPAQL 60
Db 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFEVERIEHTKDDGSVMTLPDPAQL 60
Qy 61 VQSEGPVAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVNMPEIPOEADD 120
Db 61 VQSEGPVAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVNMPEIPOEADD 120
Qy 121 GGLNFGSGFTISKNGYILTNTHVVAAGMSIVLNDKREYTAKLIGSDVQSVALLKIDA 180
Db 121 GGLNFGSGFTISKNGYILTNTHVVAAGMSIVLNDKREYTAKLIGSDVQSVALLKIDA 180
Qy 181 TEELPVKIKGNPKULKPGEMVAAICAPFGFDSVTAGIVSAKGRSLPESYTPFIQDVA 240
Db 181 TEELPVKIKGNPKULKPGEMVAAICAPFGFDSVTAGIVSAKGRSLPESYTPFIQDVA 240
Qy 241 INPNSGGPLFNLKQVVGINSQIYSRSQGFMSFALPIDVAMNVABQLKNTGKYVRGQ 300
Db 241 INPNSGGPLFNLKQVVGINSQIYSRSQGFMSFALPIDVAMNVABQLKNTGKYVRGQ 300
Qy 301 LGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIYLSLDGGEIRSSGDL 360
Db 301 LGVIIQEVSYGLAOSFGLDKAGALIAKILPGSPAERAGLRAGDIYLSLDGGEIRSSGDL 360
Qy 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOGSTFSVES 420
Db 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOGSTFSVES 420
Qy 421 AGITLQHTDSSGKHLVVVRVSDPAERAGLRHGEDEILAVRASP 463
Db 421 AGITLQHTDSSGKHLVVVRVSDPAERAGLRHGEDEILAVRASP 463

```

```
RESULT 7
AAVS2994
ID AAVS2994 standard; Protein: 499 AA.
XX
AC AAVS2994;
XX
DT 21-FEB-2000 (first entry)
XX
DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.
XX
KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
KM antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
XX invasive bacterial disease; antibacterial.
XX
OS Neisseria meningitidis.
XX
PN WO955872-A1.
XX
04-NOV-1999.
XX
20-APR-1999; 99WO-EP02765.
XX
23-APR-1998; 98GB-0008734.
XX
(SM1K ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
Ruelle J;
XX
WPI; 2000-052809/04.
XX
DR N-PSDB; AA233306.
XX
PT Novel polynucleotides and polypeptides from Neisseria meningitis used
XX to prepare vaccines against bacterial infections
XX
Claim 3; Page 77-78; 94pp; English.
XX
XX
XX The present sequence represents a BASB013 polypeptide isolated from
XX Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be
XX employed as research reagents and material for the discovery of
XX treatments and diagnostics for diseases, particularly human diseases.
XX They can be used for diagnosis of disease, staging of disease, or
XX determining response of an infectious organism to drugs. The
XX polynucleotides may be used as a source for hybridisation probes, and
XX for screening of genetic mutations, serotype, organism or strain
XX identification. Identification of mutation in BASB013 sequences, and as
XX components of arrays which are useful for diagnostic and prognostic
XX purposes. The polypeptides can be used to produce antibodies. The
XX polypeptides can also be used in vaccine formulations, and to identify
XX agonists and antagonists. The polypeptides, antibodies, agonists and
XX antagonists (which are bacteriostatic) are used for the treatment and
XX prevention of diseases such as upper respiratory tract infection,
XX invasive bacterial diseases such as bacteraemia and meningitis, and for
XX the development and screening of antibacterial drugs. They are also used
XX in the prevention of adhesion of bacteria to eukaryotic matrix proteins
XX on in-dwelling devices, or to extracellular proteins on wounds, and to
XX thus prevent tissue damage and/or block the normal progression of
XX pathogenesis in infections initiated other than by the implantation of
XX in-dwelling devices or by other surgical techniques.
XX
SQ Sequence 499 AA;
XX
Query Match 93.6%; Score 2202; DB 21; Length 499;
Best Local Similarity 94.0%; Pred. No. 1,1e-177;
Matches 435; Conservative 10; Mismatches 18; Indels 0; Gaps 0;
XX
XX 1 VFKKYVYFALAAACALLACCEKAGSFPGADKKEASFYERIEHKKDQSVSMLPDPFQQL 60
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1 MKKKYIYLALALCAASLACGDKRAGSFGADKKEASFYERIEHKKDQSVSMLPDPFQQL 60
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
XX 61 VQSEGPVVNIQAPAPRTONGSGNAETDSDPLADSDPFYEFPRQLVPMPEIQEADAD 120
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
DB 61 VQSEGPVVNIQAPAPRTONGSGNAETDSDPLADSDPFYEFPRQLVPMPEIQEADAD 120
QY GGLNFGSGFIISKNGYIILNTHVAVAGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDA 180
DB GGLNFGSGFIISKNGYIILNTHVAVAGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDA 180
QY TEELPVVKIGNPKNIKPEBWAALGAPFGFDSVTAGIVSAKGRSLPNESYTFPIQTDVA 240
DB TEELPVVKIGNPKNIKPEBWAALGAPFGFDSVTAGIVSAKGRSLPNESYTFPIQTDVA 240
QY INPGNSGGPLFNLKQOVGINQIYRSRGFNGISPAIPIDVAMNVAEOLKNTGXQORQ 300
DB INPGNSGGPLFNLKQOVGINQIYRSRGFNGISPAIPIDVAMNVAEOLKNTGXQORQ 300
QY LGVIIQEVSYGLAQSFGLDKASGALIAKILPSPPARAGLQAGDYLISLDCGEIRSSGDL 360
DB LGVIIQEVSYGLAQSFGLDKASGALIAKILPSPPARAGLQAGDYLISLDCGEIRSSGDL 360
QY FVMVGAIITPGKEVSLGVWRKGEITIKAKLGNAEHTGASSTDEAPYTEQSGTFSVES 420
DB FVMVGAIITPGKEVSLGVWRKGEITIKAKLGNAEHTGASSTDEAPYTEQSGTFSVES 420
QY AGITLQTHTDSSGKLVVVRVSDAERAGLRGDEILAVRASP 463
DB AGITLQTHTDSSGKLVVVRVSDAERAGLRGDEILAVRASP 463
RESULT 8
AAVS2993
ID AAVS2993 standard; Protein: 499 AA.
XX
AC AAVS2993;
XX
DT 21-FEB-2000 (first entry)
XX
DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.
XX
XX
XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
XX antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
XX invasive bacterial disease; antibacterial.
XX
OS Neisseria meningitidis.
XX
PN WO955872-A1.
XX
04-NOV-1999.
XX
20-APR-1999; 99WO-EP02765.
XX
23-APR-1998; 98GB-0008734.
XX
(SM1K ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
Ruelle J;
XX
WPI; 2000-052809/04.
XX
DR N-PSDB; AA233305.
XX
PT Novel polynucleotides and polypeptides from Neisseria meningitis used
XX to prepare vaccines against bacterial infections
XX
Claim 5; Page 74-76; 94pp; English.
XX
XX
XX The present sequence represents a BASB013 polypeptide isolated from
XX Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be
XX employed as research reagents and material for the discovery of
XX treatments and diagnostics for diseases, particularly human diseases.
XX They can be used for diagnosis of disease, staging of disease, or
XX determining response of an infectious organism to drugs. The
XX polynucleotides may be used as a source for hybridisation probes, and
XX for screening of genetic mutations, serotype, organism or strain
XX identification. Identification of mutation in BASB013 sequences, and as
XX components of arrays which are useful for diagnostic and prognostic
```

CC purposes. The polypeptides can be used to produce antibodies. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC agonists and antagonists. The polypeptides, antibodies, agonists and
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases such as upper respiratory tract infection,
 CC invasive bacterial diseases such as bacteraemia and meningitis, and for
 CC the development and screening of antibacterial drugs. They are also used
 CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins
 CC on in-dwelling devices, or to extracellular proteins on wounds, and to
 CC thus prevent tissue damage and/or block the normal progression of
 CC pathogenesis in infections initiated other than by the implantation of
 CC in-dwelling devices or by other surgical techniques.

XX Sequence 499 AA:

Query Match 93.4%; Score 2197; DB 21; Length 499;
 Best Local Similarity 93.7%; Pred. No. 2.9e-177;
 Matches 434; Conservative 10; Mismatches 19; Indels 0; Gaps 0;

```

1 VFKKYQYALALCALLAGCEKASFGADKKEASFVERIHTKDDGSVSMLLPDPQL 60
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MFKKYQYALALCAASLAGCDKASFGADKKEASFVERIHTKDDGSVSMLLPDPQL 60
VOSGPAVVNIQAAPARTONGSNAETDSDPLADSPFYEFFKRLVPMPEIPOBEAD 120
61 VOSGPAVVNIQAAPARTONGSNAETDSDPLADSPFYEFFKRLVPMPEIPOBEAD 120
61 VOSGPAVVNIQAAPARTONGSNAETDSDPLADSPFYEFFKRLVPMPEIPOBEAD 120
121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYTKLIGSDVQSVALLKIDA 180
121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYTKLIGSDVQSVALLKIDA 180
121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYTKLIGSDVQSVALLKIDA 180
181 TEELPVVKIGNPKXNKPGEVVAATGAPGPDNSVTAGISAGRSLPNPSYPIQTDVA 240
181 TEELPVVKIGNPKXNKPGEVVAATGAPGPDNSVTAGISAGRSLPNPSYPIQTDVA 240
181 TEELPVVKIGNPKXNKPGEVVAATGAPGPDNSVTAGISAGRSLPNPSYPIQTDVA 240
241 INPGNSGGLFNLKQVVGINSQIYSRSGFMGISPAIPIDVAMNVAEQLKTKGVQRCQ 300
241 INPGNSGGLFNLKQVVGINSQIYSRSGFMGISPAIPIDVAMNVAEQLKTKGVQRCQ 300
241 INPGNSGGLFNLKQVVGINSQIYSRSGFMGISPAIPIDVAMNVAEQLKTKGVQRCQ 300
301 LGVITIOEVSYGLAQSGFLDKASGALLAKILPGSPARRAGIQAGDIYLSLDGGRIRSSGDL 360
301 LGVITIOEVSYGLAQSGFLDKASGALLAKILPGSPARRAGIQAGDIYLSLDGGRIRSSGDL 360
301 LGVITIOEVSYGLAQSGFLDKASGALLAKILPGSPARRAGIQAGDIYLSLDGGRIRSSGDL 360
361 PVMVGAITGKEKESLVGWRKGEKITIVAVKLGNASBOTGSSSEPDKAPYAEHQSGTISVES 420
361 PVMVGAITGKEKESLVGWRKGEKITIVAVKLGNASBOTGSSSEPDKAPYAEHQSGTISVES 420
421 AGITLQTHTDSGSKHLVVVRVSDAERAGLRHGDDEILAVRASP 463
421 AGITLQTHTDSGSKHLVVVRVSDAERAGLRHGDDEILAVRASP 463
421 AGITLQTHTDSGSKHLVVVRVSDAERAGLRHGDDEILAVRASP 463

```

RESULT 9
 ID AAY70414 standard; Protein; 475 AA.
 AAY70414;

03-JUL-2000 (first entry)

Neisseria meningitidis NMASP protein-3.

NMASP: non-cytosolic; antibacterial; antiinflammatory; cytotoxic;
 anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;
 Neisserial infection; meningitidis; septicemia.

Neisseria meningitidis.

Key Location/Qualifiers

FT Misc-difference 19 /note= "Encoded by GAA"
 FT Misc-difference 35 /note= "Encoded by GCC"

FT Misc-difference 61 /note= "Encoded by GGC"
 FT Misc-difference 111 /note= "Encoded by AAC"
 FT Misc-difference 122 /note= "Encoded by GCC"
 FT Misc-difference 171 /note= "Encoded by GGC"
 FT Misc-difference 171 /note= "Encoded by AAT"
 FT Misc-difference 194 /note= "Encoded by GGCATC"
 FT Misc-difference 298 /note= "Encoded by AGC"
 FT Misc-difference 317 /note= "Encoded by AGC"
 FT Misc-difference 364 /note= "Encoded by AGC"
 FT Misc-difference 373 /note= "Encoded by GCC"
 FT Misc-difference 410 /note= "Encoded by ACC"
 FT Misc-difference 428 /note= "Encoded by AAA"
 FT Misc-difference 436 /note= "Encoded by CAC"
 FT Misc-difference 437 /note= "Encoded by AGG"
 FT Misc-difference 438 /note= "Encoded by GCA"
 FT Misc-difference 440.475 /note= "Encoded by AGT"
 FT Misc-difference 440.475 /note= "Encoded by CGTCAA"

WO200012535-A2.

09-MAR-2000.

01-SEP-1999; 99WO-US19663.

01-SEP-1998; 98US-0098685.

(ANTE-) ANTEX BIOLOGICS INC.

Jackson WJ, Harris AM;

WPI; 2000-256581/22.

N-PSDB; AA251539.

Neisseria meningitidis NMASP polypeptide, nucleotide sequences and
 antibodies, useful in vaccines against infection -

Claim 5; Page 71-73; 75pp; English.

The present sequence is the Neisseria meningitidis NMASP protein.
 NMASP is a non-cytosolic protein, with antibacterial and
 antiinflammatory activity. It shows sequence similarity to E. coli
 Degr (HcrA) protein. NMASP proteins can be used as ligands to detect
 antibodies elicited in response to N. meningitidis infections. Cytotoxic
 anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA
 may be used for diagnosis, therapy or prophylaxis of Neisserial
 infections such as, bacterial meningitidis and septicemia.

Sequence 475 AA;

Query Match 90.5%; Score 2130.5; DB 21; Length 475;
 Best Local Similarity 96.4%; Pred. No. 1.2e-171;
 Matches 424; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

```

24 AGSPFGADKKEASFVERIHTKDDGSVSMLLPDPALVOSGPAVVNIQAAPARTONGS 83
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 AGSPFGADKKEASFVERIHTKDDGSVSMLLPDPALVOSGPAVVNIQAAPARTONGS 60
GNAETDSDPLADSPFYEFFKRLVPMPEIPOBEADGGLNFGSGFIISKNGYILTNTHV 143
84 GNAETDSDPLADSPFYEFFKRLVPMPEIPOBEADGGLNFGSGFIISKNGYILTNTHV 143
61 SNAETDSDPLADSPFYEFFKRLVPMPEIPOBEADGGLNFGSGFIISKNGYILTNTHV 120

```


QY 144 VAGMGSIKVLNDRKREYTAKLIGSDVQSDVALLKIDATELPPVKIGNPKLKEGEWAA 203
 DB 121 VTGMSGIKVLNDRKREYTAKLIGSDVQSDVALLKIDATELPPVKIGNPKLKEGEWAA 180
 QY 204 ICAPFGFDSVTVGIYSAKGRSLPNESYTPPIQTDVAINPNSGGLPFLNKGOVVGINSQ 263
 DB 181 IGAPFGFDSVTVAG-VSAKGRSLPNESYTPPIQTDVAINPNSGGLPFLNKGOVVGINSQ 239
 QY 264 IYRSRGFGMGISPAIPFDVAMNVAEQLKNTGKVRGGLGVIOEVSYGIAQSPGLDKASG 323
 DB 240 IYRSRGFGMGISPAIPFDVAMNVAEQLKNTGKVRGGLGVIOEVSYGIAQSPGLDKASG 299
 QY 324 ALIAXILPGSPARAGLQAGDIVLSLDGGEIRSSGDLPMVVGATTPGKEVSLGYWRKGE 383
 DB 300 ALIAXILPGSPARAGLQAGDIVLSLDGGEIRSSGDLPMVVGATTPGKEVSLGYWRKGE 359
 QY 384 ITTKAKLGNAAEHTGASSKTDEAPYTEQSGTFSVESAGITLQTHTDSGKHLVVRVSD 443
 DB 360 ITTKAKLGNAAEHTGASSKTDEAPYTEQSGTFSVESAGITLQTHTDSGKHLVVRVSD 419
 444 AABRAGLRHGDDELAVRASP 463
 DB 420 AABRAGLRHGDDELAVRASP 439

RESULT 10

AAY83151
 ID AAY83151 standard; Protein; 414 AA.

AC AAY83151;
 DT 24-JUL-2000 (first entry)

NGSP polypeptide of *Neisseria gonorrhoeae*.

NGSP, polypeptide; peptide; vaccine; immune response; antibody;
 cellular matrix; adherence domain; ligand; detection; diagnosis;
 screening; probe; primer; prophylaxis; therapy.

OS *Neisseria gonorrhoeae*.

PN WO200012133-A1.

PD 09-MAR-2000.

PF 01-SEP-1999; 99WO-US20070.

PY 01-SEP-1998; 98US-0098685.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Jackson WJ, Harris AM;

WPI; 2000-237782/20.

N-PSDB; AA293415.

Non-cytosolic NGSP polypeptide and polynucleotide sequence from
Neisseria useful for diagnosis, prevention or treatment of *Neisseria*
 infections

Claim 5; Page 63-64; 68pp; English.

The NGSP polypeptide of *N. gonorrhoeae* has conserved Arg-Gly-Asp and
 Arg-Gly-Asn motifs near the C-terminus which function as adherence
 domains for extracellular matrix proteins. Using the NGSP polypeptide
 as a vaccine produces antibodies which inhibit binding of *N.*
gonorrhoeae to the host's cellular matrix reducing attachment and/or
 subsequent invasion. The NGSP polypeptide and its peptide fragments
 can be used to immunize an animal and produce an immune response.
 They can also be used as ligands to detect antibodies elicited in
 response to *Neisseria* infections and also as antigens or immunogens
 for inducing *Neisseria*-specific antibodies which are useful in

immunosays to detect *Neisseria* in biological specimens. Nucleotides
 encoding NGSP or its fragments can be used as probes to identify
Neisseria in biological specimens by hybridization or polymerase
 chain reaction amplification. The NGSP polypeptide can also be used
 in screening assays to identify agents and compounds which useful as
 diagnostic, prophylactic or therapeutic agents against *Neisseria*
 infection.

SQ Sequence 414 AA;

Query Match 88.9%; Score 2092; DB 21; Length 414;

Best Local Similarity 100.0%; Pred. No. 1,7e-168;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 MLPPFAQLVQSEGAVVNIQAAPARTQNGSNAETBDDPLADSPFEFFKRLVPPNP 111
 DB 1 MLPPFAQLVQSEGAVVNIQAAPARTQNGSNAETBDDPLADSPFEFFKRLVPPNP 60
 QY 112 EIPQEAADGGINFSGFIIISKNGYILNTTHVVAAGMSIKVLLNDRKREYTAKLIGSDVQ 171
 DB 61 EIPQEAADGGINFSGFIIISKNGYILNTTHVVAAGMSIKVLLNDRKREYTAKLIGSDVQ 120
 QY 172 DVALLKIDATELPPVKIGNPKLKEGEWAAIGAPFGFDSVTVGIYSAKGRSLPNESY 231
 DB 121 DVALLKIDATELPPVKIGNPKLKEGEWAAIGAPFGFDSVTVGIYSAKGRSLPNESY 180
 QY 232 TPIQTDVAINPNSGGLPFLNKGOVVGINSQIYRSRGFGMGISPAIPFDVAMNVAEQLK 291
 DB 181 TPIQTDVAINPNSGGLPFLNKGOVVGINSQIYRSRGFGMGISPAIPFDVAMNVAEQLK 240
 QY 292 NTGKVRGQLGVIOEVSYGIAQSPGLDKASGALIAKILPGSPARAGLQAGDIVLSLDG 351
 DB 241 NTGKVRGQLGVIOEVSYGIAQSPGLDKASGALIAKILPGSPARAGLQAGDIVLSLDG 300
 QY 352 GEIRSSGGLPMVVGATTPGKEVSLGYWRKGEITTKAKLGNAAEHTGASSKTDEAPYTRQ 411
 DB 301 GEIRSSGGLPMVVGATTPGKEVSLGYWRKGEITTKAKLGNAAEHTGASSKTDEAPYTRQ 360
 QY 412 QSGTFSVESAGITLQTHTDSGKHLVVRVSDAABRAGLRHGDDELAVRASP 465
 DB 361 QSGTFSVESAGITLQTHTDSGKHLVVRVSDAABRAGLRHGDDELAVRASP 414

RESULT 11

AAY70409
 ID AAY70409 standard; Protein; 448 AA.

AC AAY70409;

DT 03-JUL-2000 (first entry)

DE *Neisseria meningitidis* NMA SP protein-1.

NMA SP, non-cytosolic; antibacterial; antiinflammatory; cytotoxic;

anti-NMA SP antibody; vaccine; diagnosis; therapy; prophylaxis;

Neisseria infection; meningitidis; septicemia.

OS *Neisseria meningitidis*.

Key Location/Qualifiers

FT Misc-difference 167 /note= "Encoded by NTC"

PN WO200012535-A2.

PD 09-MAR-2000.

PF 01-SEP-1999; 99WO-US19663.

PY 01-SEP-1998; 98US-0098685.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Jackson WJ, Harris AM;
 XX WPI; 2000-256581/22.
 DR N-PSDB; AA251533.
 XX
 PT Neisseria meningitidis NMASP polypeptide, nucleotide sequences and
 XX antibodies, useful in vaccines against infection
 XX
 PS Claim 5; Page-; 75pp; English.
 CC The present sequence is the Neisseria meningitidis NMASP protein.
 CC NMASP is a non-cytosolic protein, with antibacterial and
 CC anti-inflammatory activity. It shows sequence similarity to E. coli
 CC DegP (HtrA) protein. NMASP proteins can be used as ligands to detect
 CC antibodies elicited in response to N. meningitidis infections. Cytotoxic
 CC anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA
 CC may be used for diagnosis, therapy or prophylaxis of Neisserial
 CC infections such as, bacterial meningitidis and septicemia.
 CC Note: The protein represented in SEQ ID NO:2 of the specification is
 CC erroneous. The present sequence is the decoded version of the nucleotide
 CC represented in AA251533.

SO Sequence 448 AA;

Query Match 85.1%; Score 2002; DB 21; Length 448;
 Best Local Similarity 96.4%; Pred. No. 8e-16i;
 Matches 397; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 52 MLDPFAVLVQSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNP 111
 DB 1 MLDPFVLVQSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNP 60
 QY 112 EIPOEADGGINFGSGFIISKNGYILTNTHVAVAGMSIKVLNDRKREYAKLIGSDVQS 171
 DB 61 EIPOEADGGINFGSGFIISKNGYILTNTHVAVAGMSIKVLNDRKREYAKLIGSDVQS 120
 QY 172 DVALLKIDATEELPVYKIGNPNKLRGEWVAAGAPFGFDSVTAGIVSAKGRSLPNESY 231
 DB 121 DVALLKIDATEELPVYKIGNPNKLRGEWVAAGAPFGFDSVTAGIVSAKGRSLPNESY 180
 QY 232 TPFITDVAINPNSGGPLFNLKGVVGINSOIYSSGGFMGISPAIPDIVANVAQLK 291
 DB 181 TPFITDVAINPNSGGPLFNLKGVVGINSOIYSSGGFMGISPAIPDIVANVAQLK 240
 QY 292 NTGKVGQGLGVYIQEVSYGLAQSGFLDKASGALIKILPGSPAERAGQAQGIIVSLSDG 351
 DB 241 NTGKVGQGLGVYIQEVSYGLAQSGFLDKASGALIKILPGSPAERAGQAQGIIVSLSDG 300
 DB 352 GEIRSSGDLPVWVGAIITPGEKESVLGWRKGEETITIRAKLGNAAEHTGASSKTDEAPYTEQ 411
 DB 301 GEIRSSGDLPVWVGAIITPGEKESVLGWRKGEETITIRAKLGNAAEHTGASSKTDEAPYTEQ 360
 QY 412 QSGTFSVESAGITLQTHDSSGKHLVYVSDAERAGLPHGDELLAVRASP 463
 DB 361 QSGTFSVESAGITLQTHDSSGKHLVYVSDAERAGLPHGDELLAVRASP 412

RESULT 12

AAVS2996
 ID AAVS2996 standard; Protein; 370 AA.

XX AAVS2996;
 XX
 DT 21-FEB-2000 (first entry)
 DE Neisseria meningitidis BASB013-C protein sequence.
 XX
 KM Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
 KM antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
 KM invasive bacterial disease; antibacterial.
 XX
 OS Neisseria meningitidis.

PN WO955872-A1.
 XX
 XX 04-NOV-1999.
 XX
 PF 20-APR-1999; 99MO-EP02765.
 XX
 PR 23-APR-1998; 98GB-0008734.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Ruelle J;
 XX
 DR WPI; 2000-052809/04.
 DR N-PSDB; AA233308.
 XX

PT Novel polynucleotides and polypeptides from Neisseria meningitidis used
 XX to prepare vaccines against bacterial infections
 XX
 PS Example 2; Page 82-83; 94pp; English.

CC The present sequence is a conserved BASB013-C polypeptide isolated
 CC from Neisseria meningitidis. BASB013 polynucleotides and polypeptides
 CC may be employed as research reagents and material for the discovery of
 CC treatments and diagnostics for diseases, particularly human diseases.
 CC They can be used for diagnosis of disease, staging of disease, or
 CC determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and
 CC for screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC agonists and antagonists. The polypeptides, antibodies, agonists and
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases such as upper respiratory tract infection,
 CC invasive bacterial diseases such as bacteraemia and meningitis, and for
 CC the development and screening of antibacterial drugs. They are also used
 CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins
 CC on in-dwelling devices, or to extracellular proteins on wounds, and to
 CC thus prevent tissue damage and/or block the normal progression of
 CC pathogens in infections initiated other than by the implantation of
 CC in-dwelling devices or by other surgical techniques.

SO Sequence 370 AA;

Query Match 77.9%; Score 1832; DB 21; Length 370;
 Best Local Similarity 97.0%; Pred. No. 1.5e-146;
 Matches 359; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 VFKKYQFEALAAALCAGLAGEKASFFGADKKEASFYERIEHTKDDGSVSLPDPFQ 60
 DB 1 VFKKYQFEALAAALCAGLAGEKASFFGADKKEASFYERIEHTKDDGSVSLPDPFQ 60
 QY 61 VOSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNPPIPOEADD 120
 DB 61 VOSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNPPIPOEADD 120
 QY 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLNDRKREYAKLIGSDVQSVALLKIDA 180
 DB 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLNDRKREYAKLIGSDVQSVALLKIDA 180
 QY 181 TPELPVYKIGNPNKLRGEWVAAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 DB 181 TPELPVYKIGNPNKLRGEWVAAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 QY 241 INPNSGGPLFNLKGVVGINSOIYSSGGFMGISPAIPDIVANVAQLKNTGKVGQ 300
 DB 241 INPNSGGPLFNLKGVVGINSOIYSSGGFMGISPAIPDIVANVAQLKNTGKVGQ 300
 QY 301 LGVYIQEVSYGLAQSGFLDKASGALIKILPGSPAERAGLQAQGIIVSLDGSIRSSGDL 360
 DB 301 LGVYIQEVSYGLAQSGFLDKASGALIKILPGSPAERAGLQAQGIIVSLDGSIRSSGDL 360

```
OY 361 PWMGAIITPG 370
DB 361 PWMGAIITPG 370
RESULT 13
AAV29294
ID AAV29294 standard; Protein: 460 AA.
AC AAV29294;
XX 25-OCT-1999 (first entry)
XX
XX Protein encoded by the PA14 degt gene.
XX
XX Human pathogen; virulence polypeptide; virulence factor;
XX pathogenic infection; Pseudomonas aeruginosa infection.
XX Pseudomonas aeruginosa.
XX
XX MO9927129-A1.
XX
XX 03-JUN-1999.
XX
XX 25-NOV-1998; 98WO-US25247.
XX
XX 25-NOV-1997; 97US-006517.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
XX Rahme LG, Tan M, Tsongalis J;
XX WPI: 1999-357851/30.
XX
XX Virulence factors useful in developing disease treatments
XX
XX Disclosure; Fig 28; 228pp; English.
XX
XX The present sequence represents a Pseudomonas aeruginosa polypeptide
XX sequence. P. aeruginosa is an opportunistic human pathogen present in
XX soil water and plants. The specification describes virulence polypeptides
XX and nucleic acid sequence encoding such polypeptides. These sequences
XX can be used to identify a compound which is capable of decreasing the
XX expression of a pathogenic virulence factor. Compounds that inhibit
XX the expression or activity of virulence factor polypeptides can be
XX used to treat pathogenic infections, especially where the infection
XX is a P. aeruginosa infection.
XX note: the sequences given in the specification were poorly legible, and
XX in some instances assumptions were made as to the identity of the
XX residue; it is therefore possible that the sequence given below is
XX not entirely correct.
XX
XX Sequence 460 AA;
XX
XX Query Match 40.9%; Score 962.5; DB 20; Length 460;
XX Best Local Similarity 51.8%; Pred. No. 7.9e-73;
XX Matches 217; Conservative 56; Mismatches 113; Indels 33; Gaps 11;
OY 54 LPDPAQVOSGEPVAVNI---QAAPAPRTONGSGNAETSDPLADSDP-FYEFKRLVFN 109
DB 27 LPDFPLVQASPAVNAVNISTRQKLP---DRAMARQGLSIPDLGLPMPFRDFLERTIPQ 82
OY 110 MEIIP---OEADDDGLNFGSGFLLSKNGYILTNTHVAVAGSTKVLNDRKREYAKLIG 166
DB 83 VFRMRPGQGRQEAQ---SLGGGFTISMDGYILTNTHVVAADDELIVRLSDSEHRAKLVG 138
OY 167 SDVQSDVALIKDATEELPVVKIGNPKNLKPGEVVAIGAPFGFDSNSTAGIVSAKGRSL 226
DB 139 ADPRSDVAVLKIEA-KNLPTLKIGDSMKLKGKENVLAIGSPFGDHSHTTAKIVSAKGRSL 197
OY 227 PNESTPTPIQTDVAINPGNSGGPLFNLKQGVVGINSGIYSRSGFGMGSFAIPIDVAMNV 286
```

```
DB 198 PMESYVPFIQTDVAINPGNSGGPLNDEGEVVGINSQIFTRSGGFMGSFAIF
OY 287 AEQLKNTGKVGORGLGVIIQEVSYGLAOSFGLDKASGALIKILGPSAERAGLOAGU..
DB 258 ADQLKKAQKVSRRGLGVIIQEVNKLAEFGLDKPSGLVAVQVLDGPPAKAGLOVGDVI 317
OY 347 LSLDGEIRSSGDLPVWVGAITPGKEVSLGYWRKGEIITIRAKLGNAAEHRTGASSKTDE- 405
DB 316 LSLNGQSINESADLPHLVGMKKPGDKINLDVIRMGQRKSLMAVGNLPD-----DDEE 370
OY 406 -----APYTEQSGTSVESAGITLQTH--TDSSGKHLVYVRVSDA-ABRAGLRGDEI 456
DB 371 IASGAPGARSSRWLGVTVAADLTAEQRKSLDIQ-GVIVKEVDGPAAVIGLRPGDVI 428
RESULT 14
AA777434
ID AAR77434 standard; Protein: 503 AA.
XX
XX AAR77434;
XX
XX 19-AUG-1996 (first entry)
XX
XX Heat shock protein of Rochalimaea henselae.
XX
XX Rochalimaea henselae; cat scratch disease; bacillary angiomatosis;
XX CSD; infection; antigen; antibody; vaccine.
XX
XX Rochalimaea henselae.
XX
XX WO9531549-A1.
XX
XX 23-NOV-1995.
XX
XX 18-MAY-1995; 95WO-US06211.
XX
XX 16-SEP-1994; 94US-0307279.
XX
XX 18-MAY-1994; 94US-0245294.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Anderson BE, Regnery RL;
XX
XX WPI: 1996-010935/01.
XX
XX N-PSDB; AAT04403.
XX
XX Nucleic acids of Rochalimaea henselae and R. quintana - methods which
XX enable the identification of R.henselae, which is a causative agent
XX of both cat scratch disease and bacillary angiomatosis
XX
XX Claim 12; Page 84-86; 100pp; English.
XX
XX The nucleic acids (AAT04402, AAT04403), fragments and antibodies
XX binding to the encoded proteins (AAR77433, AAR77434), may be used in the
XX diagnosis and detection of cat scratch disease (CSD) and bacillary
XX angiomatosis caused by R. henselae. The proteins or fragments of
XX them, may be used in vaccines to protect against R. henselae
XX infection.
XX
XX Sequence 503 AA;
XX
XX Query Match 31.1%; Score 731.5; DB 17; Length 503;
XX Best Local Similarity 39.5%; Pred. No. 3.2e-53;
XX Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;
OY 8 FALAALCAALLACCEKAGSFFGADKKEASFVERLEHTKDDGSVSMLLPDAQVOSGEP 67
DB 15 FSAALERTALFFSGC---GSLMTTKAHANV-----FSSLMQOGGFADIVGQKPA 62
OY 68 VVNIQAAPAPRTON---GSGNAETSDPLADSDP---FYEFKRLVYVNMMEIIPDEADD 120
DB 63 VVSVOYKSNKKKEWFFSDPFSPPGQLPDQHPLKKKFFODFYNRDKPSKXSL-QRSHRL 121
```

QY 121 GGLNGSGFIIISKNGYILNTHTHVAGMGSIKVLLANDKREYTAKLIGSDVQSVALLKIDA 180
 Db 122 RPIAFSGSFIISSDGYIVTNNHVISDGTSYAVVLDGTEINAKLIGTDPRTDLAVLKVNE 181
 QY 181 TEELPVVXIGNPKNKKPGEWVAAIGAPFGEFNSVTAGIVSAKGRSLPNESYTFPIQTDVA 240
 Db 182 KRKFSYVPGFDDSKLRVGDVVAIIGNPFLGGITVIGIVSARGRDIGTGVDYDFPIQIDAA 241
 QY 241 INPNSGGPFLFNKQGVVGINSGIYSRSGGFMGISFPAIPIDVAMNVAEQLKNTGKYQRG 300
 Db 242 VNRNGSGPTEFDLNGKVVGVNTAIFSPSGANGVIAIPAATANEVYQQLIBKGLVQRGW 301
 QY 301 LGVIGIIEFVSYGIAQFGLDKASGALIAKLIPGSPARAGLQAGDIYLSLDGGEIRSSGDL 360
 Db 302 LGVGIQPVTKETISDIGHKEAKGALITDPLKG-PAKAGIKAGDVIISVGEKINDVRDL 360
 QY 361 PVMVGAITPGKEVSLGVWRKGEETIKAKLGNAEHTGASSKTDEAPYR-EGQSGTFPSVE 419
 Db 361 AKRIANMSFGEIVTIGVWKSKEENIKVKLDSMPED--ENMKDGSKYSNEHNSDETLE 417
 QY 420 SAGITTLQHTDSSGKHLVVRV---SDAERAGLRHGDDEILAV 459
 Db 418 DYGLIVAPSDDLG--LVVTVDVDPDSDADK-GIRGDIYIVV 457

RESULT 15

AAG78605
 ID AAG78605 standard; Protein; 474 AA.

AC AAG78605;

DT 20-NOV-2001 (first entry)

DE Lawsonia intracellularis protein SEQ ID NO: 7.

KM HtrA; PonA; HycC; YefW; ABC1; Omp100; Lawsonia intracellularis infection;
 KM vaccine.

OS Lawsonia intracellularis.

PN JP2001169787-A.

PD 26-JUN-2001.

PF 20-OCT-2000; 2000JP-0320736.

PR 22-OCT-1999; 99US-0160922.

(PFIZ) PFIZER PROD INC.

DR WPI; 2001-592540/67.

PT Lawsonia intracellularis polynucleotide and encoded protein, used to
 prevent Lawsonia intracellularis infection -

PS Claim 12; Page 47-49; 67pp; Japanese.

CC The present invention provides isolated polynucleotides encoding HtrA,
 CC PonA, HycC, YefW, ABC1 or Omp100 protein of Lawsonia
 CC intracellularis. The sequences can be used in vaccines for the prevention
 CC of Lawsonia intracellularis infection. The present sequence is a protein
 CC of the invention.

XX Sequence 474 AA;

Query Match 30.0%; Score 705; DB 22; Length 474;

Best Local Similarity 39.3%; Pred. No. 5,1e-51;
 Matches 169; Conservative 67; Mismatches 152; Indels 42; Gaps 11;

QY 54 LPDFAQDVOSGPAVWNIOA-APAFRTONGSGNAGTDSPLADSDPFYE-FKRLVPPNMP 111
 Db 28 LPNRFVPLVXKASKAVVNISTEKIIPR-----GRTEFPMEMFRGLPGRFFRFPQFEPKCP 82

QY 112 --EIPOEADDDGINFSGGFIISKNGYILNTHTHVAGMGSIKVLL---NDKREYTAKL 165
 Db 83 DSQIHKQR-----SLGTGFIISDGYIVTNNHVIBGADSVRVNLEGTSGKEESLPAYI 136
 QY 166 GSDVQSVALLKIDATEELPVVKIGNPKNKKPGEWVAAIGAPFGEFNSVTAGIVSAKGRS 225
 Db 137 GRDEETDLALLKVKSKDSLPLYLIFGNSDTEVEGEMVLAIGNPFGHTVGTGILSAKGRD 196
 QY 226 LPNESYTFPIQTDVAINPNSGGPFLFNKQGVVGINSGIYSRSGGFMGISFPAIPIDVAMN 285
 Db 197 IHAGPFNFQITDASINPNSGGPFLINSGQVGINTRIMA-SG--QGIQFAPSSMADR 253
 QY 286 VAEQLKNTGKYQRQGLVIGIIEFVSYGIAQFGLDKASGALIAKLIPGSPARAGLQAGDI 345
 Db 254 IIEQLKTNKKVSRGWLIGVTIQDVDTNTAKALGLSQAKGALVGSVPPDPADKAGLKVGI 313
 QY 346 VLSLDGGEIRSSGDLPMVGAITPGKEVSLGVWRKGEETIKAKLGNAEHTGASSKTDE 405
 Db 314 VTQADGKQIDSSASLLKALIAIKPFPFSVVKLKWVRDGSKDISITLGERKTTSSQKQSSPE 373
 QY 406 A-----PYTEOQSGTFPSVESAGITTLQHTDSSGKHLVVRVSDAERAGLRHGD 455
 Db 374 SLPGALGLSVRPLTQESKSFVVK-LGI-----GLVVSVEPPNKPASEAGIRBODI 423
 QY 456 ILAVRASPRQ 465
 Db 424 ILSANLKPLQ 433

Search completed: July 11, 2003, 10:31:20
 Job time : 75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:30:07 ; Search time 27 Seconds
(without alignments)
506.728 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKKYQIPALALCALIAG.....ERAGLRHDEILAVASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

1 number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990.5	42.1	460	4 US-09-199-637A-132	Sequence 132, App
2	731.5	31.1	503	1 US-08-245-294-8	Sequence 8, Appl
3	731.5	31.1	503	1 US-08-474-499-8	Sequence 8, Appl
4	731.5	31.1	503	4 US-08-307-279A-8	Sequence 8, Appl
5	731.5	31.1	503	4 US-09-525-310-8	Sequence 8, Appl
6	731.5	31.1	503	5 PCT-US95-06211-8	Sequence 8, Appl
7	646	27.5	463	1 US-08-485-569-2	Sequence 2, Appl
8	646	27.5	463	1 US-08-480-993-2	Sequence 2, Appl
9	646	27.5	463	1 US-07-903-079B-2	Sequence 2, Appl
10	633.5	26.9	475	1 US-08-350-741-2	Sequence 2, Appl
11	633.5	26.9	475	1 US-08-463-875A-2	Sequence 2, Appl
12	632	26.9	463	1 US-08-278-091-2	Sequence 2, Appl
13	632	26.9	463	1 US-08-483-859-2	Sequence 2, Appl
14	632	26.9	463	1 US-08-472-173-2	Sequence 2, Appl
15	632	26.9	463	1 US-08-487-167-2	Sequence 2, Appl
16	632	26.9	463	2 US-08-482-816-2	Sequence 2, Appl
17	632	26.9	463	2 US-08-296-149-2	Sequence 2, Appl
18	632	26.9	463	2 US-08-801-499-2	Sequence 2, Appl
19	632	26.9	463	2 US-08-615-271-2	Sequence 2, Appl
20	632	26.9	463	3 US-09-074-660-2	Sequence 2, Appl
21	632	26.9	463	3 US-09-074-659-2	Sequence 2, Appl
22	632	26.9	463	3 US-09-106-468-2	Sequence 2, Appl
23	632	26.9	463	4 US-09-106-466A-2	Sequence 2, Appl
24	632	26.9	463	4 US-09-106-467-2	Sequence 2, Appl
25	631.5	26.8	475	1 US-08-278-091-6	Sequence 6, Appl
26	631.5	26.8	475	1 US-08-483-859-6	Sequence 6, Appl
27	631.5	26.8	475	1 US-08-472-173-6	Sequence 6, Appl

28	631.5	26.8	475	2 US-08-487-167-6	Sequence 6, Appl
29	631.5	26.8	475	2 US-08-482-816-6	Sequence 6, Appl
30	631.5	26.8	475	2 US-08-296-149-6	Sequence 6, Appl
31	631.5	26.8	475	2 US-08-801-499-6	Sequence 6, Appl
32	631.5	26.8	475	2 US-08-615-271-6	Sequence 6, Appl
33	631.5	26.8	475	3 US-09-074-660-6	Sequence 6, Appl
34	631.5	26.8	475	3 US-09-074-659-6	Sequence 6, Appl
35	631.5	26.8	475	3 US-09-106-468-6	Sequence 6, Appl
36	631.5	26.8	475	4 US-09-106-466A-6	Sequence 6, Appl
37	631.5	26.8	475	4 US-09-106-467-6	Sequence 6, Appl
38	606.5	25.8	472	1 US-08-278-091-5	Sequence 5, Appl
39	606.5	25.8	472	1 US-08-483-859-5	Sequence 5, Appl
40	606.5	25.8	472	1 US-08-472-173-5	Sequence 5, Appl
41	606.5	25.8	472	2 US-08-487-167-5	Sequence 5, Appl
42	606.5	25.8	472	2 US-08-482-816-5	Sequence 5, Appl
43	606.5	25.8	472	2 US-08-296-149-5	Sequence 5, Appl
44	606.5	25.8	472	2 US-08-801-499-5	Sequence 5, Appl
45	606.5	25.8	472	2 US-08-615-271-5	Sequence 5, Appl

ALIGNMENTS

```
RESULT 1
US-09-199-637A-132
; Sequence 132, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Cao, Hui
; APPLICANT: Drenthard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199, 637A
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066, 517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 132
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-132
Query Match 42.1%; Score 990.5; DB 4; Length 460;
Best Local Similarity 52.5%; Pred. No. 4,4e-84;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;
54 LDPFAVLVSEGPVAVNI---QAAPARTONGSNAETSDPLADSP-PEFPKRLVFN 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
27 LDPFTLVQASAVVNISTRKLP---DRAARGLSLPDLGLPMPRDFERTIPQ 82
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110 MEIIP---QEADGGLNFGSGFTISKNGYILTNTHVVAAGSIKVLANDKREYAKLIG 166
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
83 VRNPRGQGRQAQ-----SLGSGFTISNDGYILTNHVVADADEILVLSRSEHKAKLVG 138
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 SVQSVVALTKLDATELELPVVKIGNKNLKRGSVVAALGAPFGEDNSVTAGIVAKGRSL 226
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 ADPRSVAAVAKIEA-KNLPTLKLGSNKLKVGSWVLIGSPFGDHSVTAGIVAKGRSL 197
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 PMSYTPFIQTDAVAINPGNSGGLPFLNKGQVVGINSQIYSRSGFMSGISPAIPDIVAMNV 286
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 PMSYTPFIQTDAVAINPGNSGGLMLBGEVVGINSQIIFRSQGFMSLPAIPDIVALNV 257
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 AQLKNTGVVQGGQLGVITIQEVSYGIAQSGFLDKASGALITAKLPGSPARAGIQAGDIV 34
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 258 ADOLKAGKVRSGWLGVIQEVNKLAESEGLDKPSGALVAQLVEDGPAKGLQVGDVI 317
QY 347 LSLDGEIRSSGDLPMVNGAITPGKEVSLGVMRKGEEITIKAKLGNAAEHTGASSKTDE- 405
Db 318 LSLNOSINSEADLPRLHVGMKRGDKNLNIVIRNGKSKLSMAVGNLPD-----DDEE 370
QY 406 ----APYTEQSGTFSVESAGITLQTH--TDSGKHLVVVRVSDA--AERAGLRHGEI 456
Db 371 IASMGAPGARSSNRLGVTVAADLTAEQRKSLDIQG--GVVIKEVQDGPAAVIGLRGCVI 428

RESULT 2

US-08-245-294-8
; Sequence 8, Application US/08245294
; Patent No. 5644047

GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.
APPLICANT: Regnery, Russell L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: DIAGNOSING
TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA
TITLE OF INVENTION: INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/245,294

FILING DATE:

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414, 612
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

INFORMATION FOR SEQ. ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 503 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-245-294-8

Query Match 31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 7.8e-60;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCALIACGCEKAGSFFGADKKEASFYERIEHTKDDGSVSMILPDPFQVOSBGR 67
Db 15 FSAALFETALFFSGC---GSSLMTTKAHANSV-----FSSLMOQGGFADIVSQVKPA 62
QY 68 VVNIQAAPARTON---GSGNAETDSDPLADSDP---FYEFPRKLVNPMBEIPOEBADD 120
Db 63 VVSVOVKSXKKKKKEWPFSPDFSTPGFQDLPDQHLKPFQDFYNRDKFSNKSIL-QRSHRL 121
QY 121 GGLNFGSGFIISKNGYILTNTTHVAVGNGSIKVLLNDRKREYTKLIGSDVQSDVALLKIDA 180
Db 122 RRIARSGSGFISDGYITVNNHVISDGSYAIVVDDGTELNAKLIGTDPRTDLAVLKYNE 181
QY 181 TBELPVYKIGNPNKLRGEEVVAALIGAPFGPNSTYAGIVSAKGRSLPMSESTTPFIQTIVA 240
Db 182 KRKFSYVDGDDSKLRVGDVVAIGNPFGLGTAGIVSARGRDIGTVYDDFIQIDAA 241

QY 241 INPNSGGLPLNLKGVVNGINSQIYSRSRGMGISFAIPIDVANNVAQLNKTGKVRQ 300
Db 242 VNRNGSGFTFDLNGKVVGNVTAIFSPSGNVGIAFAIPATNAEVEVQQLLEKGLVORGM 301
QY 301 LGVITIOEVSYGLAOSGGLDPAKAGLAKIIPGSAERAGLOAGDIVSLDGEIRSSGTL 360
Db 302 LGVQIQPVYKEISDSIGLKEAKGALITDPLKG--PAARAGIRAGVITISVNGEKINDVRDL 360
QY 361 PVMVGAITPGKEVSLGVMRKGEEITIKAKLGNAAEHTGASSKTDEAPYT--EQSGTFSVE 419
Db 361 AKRIANMSPEETVTLGVWKGSKKEENIKYKLDSPED---ENMKDQSKTSNHEGNSDETLE 417
QY 420 SAGITLQTHDSSGKHLVVVRV---SDAERAGLRHGEILAV 459
Db 418 DGLIYVAPSDDGLG--LVVTVDVDDSDAAUK--GIRPGVITV 457

RESULT 3

US-08-474-499-8
; Sequence 8, Application US/08474499
; Patent No. 5693776

GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.
APPLICANT: Regnery, Russell L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE
TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,499
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/245,294
FILING DATE: 18-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414, 612
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

INFORMATION FOR SEQ. ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-474-499-8

Query Match 31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 7.8e-60;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCALIACGCEKAGSFFGADKKEASFYERIEHTKDDGSVSMILPDPFQVOSBGR 67
Db 15 FSAALFETALFFSGC---GSSLMTTKAHANSV-----FSSLMOQGGFADIVSQVKPA 62
QY 68 VVNIQAAPARTON---GSGNAETDSDPLADSDP---FYEFPRKLVNPMBEIPOEBADD 120

```

Db      63 VVSVOVSKNKKKKWFPSDFSTPGFDQLPDQHPKFFODFYNRDKPSNKSU--QRSHRL 121
Qy      121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKYLLANDKREYAKLIGSDVQSPVALLKIDA 180
      122 RPIAFSGSFFIISDGYITNNHVISDGTSYAVVLDGTELNKLICTDPRDLAVALKNE 181
Qy      181 TEELPVVKIKGNPKLKGEMVAALGAPFGFDNSVTAGIVSAKGRSLPNESYTPFIQTDA 240
      182 KRKFSYVDFGDDSKLRVGDWVAALGNPFGIGTGTAGIVSARGRDIGTGVDDFIQIDA 241
Qy      241 INPNSGSGPLFLKGOVVGINSQIYSRSGFMGISFAIPIDVAMNVAEQLKTKGVQGG 300
      242 VNRGNSGSPFDLNGKVGAVNTAIFSPSGNVGIAFAIPATVANEVQQLIEKGLVQRGW 301
Qy      301 LGVIOEVSYGLAOSFGIDKASGALIAKILPGSPAERAGLOAGDIVLSLDCGEIRSSGDL 360
      302 LGVQIQPTVTEISDSIGLKEAKGALITDPLKG-PAKAKGIRAGDVITISVNGEKINDVRDL 360
Qy      361 PVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGASSTDEAPYT-EOQSGTFSVE 419
      361 AKRIANMSPGETVTLGVWMSGKEENIKVTLDSMPED---ENMKDGSKYSNEHNSDETLE 417
Qy      420 SAGITLQHTDSSGKHLVVVRV---SDAERAGLRHGDILLAV 459
      418 DYGLIVAPSDDGLG--LVVTVDVDPDSDAADK-GIRPGDVIVTV 457
Db

```

RESULT 4

```

US-08-307-279A-8
; Sequence 8, Application US/08307279A
; Patent No. 5736347
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,279A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-307-279A-8

```

```

Query Match      31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 7,8e-50;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

```

```

Qy      8 FALAALCAALLAGECEKAGSFFGADKKEASFVERIEHTKDDGSVSMLLPDPFQOLVQSEGA 67
      15 FSALELTALFFSGC---GSSIMTTKHAANSV-----FSSIMQOQGFADIVSQVRKA 62
Qy      68 VVNIQAAPARTQN---GSGNAETDSPPLADSDP----FYFFPKRLVNMPEIPQBEADD 120
      63 VVSVOVSKNKKKKWFPSDFSTPGFDQLPDQHPKFFODFYNRDKPSNKSU--QRSHRL 121
Qy      121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKYLLANDKREYAKLIGSDVQSPVALLKIDA 180
      122 RPIAFSGSFFIISDGYITNNHVISDGTSYAVVLDGTELNKLICTDPRDLAVALKNE 181
Qy      181 TEELPVVKIKGNPKLKGEMVAALGAPFGFDNSVTAGIVSAKGRSLPNESYTPFIQTDA 240
      182 KRKFSYVDFGDDSKLRVGDWVAALGNPFGIGTGTAGIVSARGRDIGTGVDDFIQIDA 241
Qy      241 INPNSGSGPLFLKGOVVGINSQIYSRSGFMGISFAIPIDVAMNVAEQLKTKGVQGG 300
      242 VNRGNSGSPFDLNGKVGAVNTAIFSPSGNVGIAFAIPATVANEVQQLIEKGLVQRGW 301
Qy      301 LGVIOEVSYGLAOSFGIDKASGALIAKILPGSPAERAGLOAGDIVLSLDCGEIRSSGDL 360
      302 LGVQIQPTVTEISDSIGLKEAKGALITDPLKG-PAKAKGIRAGDVITISVNGEKINDVRDL 360
Qy      361 PVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGASSTDEAPYT-EOQSGTFSVE 419
      361 AKRIANMSPGETVTLGVWMSGKEENIKVTLDSMPED---ENMKDGSKYSNEHNSDETLE 417
Qy      420 SAGITLQHTDSSGKHLVVVRV---SDAERAGLRHGDILLAV 459
      418 DYGLIVAPSDDGLG--LVVTVDVDPDSDAADK-GIRPGDVIVTV 457
Db

```

RESULT 5

```

US-09-525-310-8
; Sequence 8, Application US/09525310
; Patent No. 6406887
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/525,310
; FILING DATE: 14-Mar-2000
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,279
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids

```


TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 US-09-525-310-8

Query Match 31.1%; Score 731.5; DB 4; Length 503;
 Best Local Similarity 39.5%; Pred. No. 7.8e-60;
 Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCAALLAGCEKAGSFFGADKKEASFVRIEHTKDQGSVSMLLPDPFAQLVQSEGRA 67
 DB 15 FSALETLFFSGC---GSSLMTTKAHANSV-----FSSLMOQOGFADIVSQVKPA 62
 QY 68 VVNIQAPAPRTQN---GSGNAETDSPLADSDP---FYEFKRLVPMNMEIPEBEADD 120
 DB 63 VVSQVQKSNKKKKEMFFSDFFSTPGFDQLPDQHPLKKFFQDFYNRDKPSNKSLL-QRSHRL 121
 QY 121 GGLNFGSGFTIISKNGYLLTNTHVVAAGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 DB 122 RPIAFSGSFFIISDGYIVTNHNVISDGTSYAVVLDGTELNKLLIGTDPRTDLAVLKVNE 181
 QY 181 TELPLPVYKGNPKLKPGEWVAALGAPFGPDNSTAGIVSAKGRSLPNESTPFIQTDVA 240
 DB 182 KRFSYVDFGDDSKLRVGDWVAALGNPFGIGTIVTLAGIVSARGRDIGTVYDDFIQIDAA 241
 QY 241 INPGNSGGLPFLNLKQGVGINSQIYSRSGGFMGIFPAIPIDVAMNVAEQLKNTGKVQRGQ 300
 DB 242 VNRGNSGGPTDLNGKVVGVNTAIFSPSGGNVGIAPAIIPAATANEVVOQLIEKGLVQRGW 301
 QY 301 LGVITIEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIVLSLDCGEIRSSGDL 360
 DB 302 LGVQIQPTKEISDSIGLKEAKGALITDPLKG--PAKAKGIRKAGDIVISVNGEKINDVRL 360
 QY 361 PVMVGAITPGKEVSLGVRKGEETITKAKLGNAAEHTGASSTDEAPYT--EQSGTSFVSE 419
 DB 361 AKRIANMSPGETVTLGVWKSCKEENIKVLDMPED--ENMKDQSKYSNEHNSDETLE 417
 QY 420 SAGITLQHTDSSGKGLVWVRV---SDAERAGLRHGEIILAV 459
 DB 418 DYGLIVAPSDDGLG--LVVTVDVDPDSDAADK-GIRPGDIVTV 457

RESULT 6
 PCT-US95-06211-8
 ; Sequence 8, Application PC/TUS9506211
 ; GENERAL INFORMATION:

APPLICANT:
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
 TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 STREET: 127 Peachtree Street, Suite 1200
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/06211
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/245,294
 FILING DATE: 18 MAY 1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Spratt, Gwendolyn D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 1414.6121

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 503 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-06211-8

Query Match 31.1%; Score 721.5; DB 5; Length 503;
 Best Local Similarity 39.5%; Pred. No. 7.8e-60;
 Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCAALLAGCEKAGSFFGADKKEASFVRIEHTKDQGSVSMLLPDPFAQLVQSEGRA 67
 DB 15 FSALETLFFSGC---GSSLMTTKAHANSV-----FSSLMOQOGFADIVSQVKPA 62
 QY 68 VVNIQAPAPRTQN---GSGNAETDSPLADSDP---FYEFKRLVPMNMEIPEBEADD 120
 DB 63 VVSQVQKSNKKKKEMFFSDFFSTPGFDQLPDQHPLKKFFQDFYNRDKPSNKSLL-QRSHRL 121
 QY 121 GGLNFGSGFTIISKNGYLLTNTHVVAAGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 DB 122 RPIAFSGSFFIISDGYIVTNHNVISDGTSYAVVLDGTELNKLLIGTDPRTDLAVLKVNE 181
 QY 181 TELPLPVYKGNPKLKPGEWVAALGAPFGPDNSTAGIVSAKGRSLPNESTPFIQTDVA 240
 DB 182 KRFSYVDFGDDSKLRVGDWVAALGNPFGIGTIVTLAGIVSARGRDIGTVYDDFIQIDAA 241
 QY 241 INPGNSGGLPFLNLKQGVGINSQIYSRSGGFMGIFPAIPIDVAMNVAEQLKNTGKVQRGQ 300
 DB 242 VNRGNSGGPTDLNGKVVGVNTAIFSPSGGNVGIAPAIIPAATANEVVOQLIEKGLVQRGW 301
 QY 301 LGVITIEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIVLSLDCGEIRSSGDL 360
 DB 302 LGVQIQPTKEISDSIGLKEAKGALITDPLKG--PAKAKGIRKAGDIVISVNGEKINDVRL 360
 QY 361 PVMVGAITPGKEVSLGVRKGEETITKAKLGNAAEHTGASSTDEAPYT--EQSGTSFVSE 419
 DB 361 AKRIANMSPGETVTLGVWKSCKEENIKVLDMPED--ENMKDQSKYSNEHNSDETLE 417
 QY 420 SAGITLQHTDSSGKGLVWVRV---SDAERAGLRHGEIILAV 459
 DB 418 DYGLIVAPSDDGLG--LVVTVDVDPDSDAADK-GIRPGDIVTV 457

RESULT 7
 US-08-485-569-2
 ; Sequence 2, Application US/08485569
 ; Patent No. 5679547
 ; GENERAL INFORMATION:
 APPLICANT: Krivan, Howard C.
 TITLE OF INVENTION: Samuel, James E.
 TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/485,569
 FILING DATE: 07-JUN-1995

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/903,079
FILING DATE: 22-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,966
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/631,698
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-049
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-569-2

```

```

Query Match 27.5%; Score 646; DB 1; Length 463;
Best Local Similarity 38.9%; Pred. No. 6,6e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;

```

```

QY 54 LPDPAQLVQSEGPAVNIQAAPARTONGSNAETDS-DPLADSDP--FYEFF-KRLVFN 109
DB LPFSVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSRSPFLDDIPEEFKFFGDRFAEQ 85
QY 110 MEIIPQEAADGGGLNFGSGFII-SKNGYILTNTHVVAAGMSIKVLLNDRKREYAKLIGSD 168
DB 86 FGGRGSKRNPRGL--GSGVTIINASKGYVLTNNHVIDGADKITVQLDGRERFAKLVGKD 143
QY 169 VQSDVALLKIDATELPELVKIGNPKNLKPGEWVAAIGAPFGFDSNVTAGIVSAKGRSLPN 228
DB 144 EQSDIALVQLEKPSNLTEIKFADSDKLAVGDFVTAIGNPFGIGQVTSIGIVSALGRSTGS 203
QY 229 ES--YTFPIQTVAIINPNSGGPLFNKQGVVINSQIYSRSGFMGISFALPIDVMNV 286
DB 204 DSGTYENYIQTDAVNRNNSGALVNLNGELIGINTALISPSGNAAGIAFAIPNSQASNL 263
QY 287 AEOLKNTKXVORGOGLVITQEVSYGLAQSFGLDKASGALIAKILPGSPAERGLQAGDIY 346
DB 264 VQQLLEFGVRRGLGIRKIGELNADLAKAFNVSAQGAFFSEVLKPSAAEKAKGLAKAGDII 323
QY 347 LSLDGEIRSSGDLPMVVGATTPGKEVSLGWRKGEETIRAKLGNAAEHTGASSTDEA 406
DB 324 TANNGQKISSFAEIRAKIATTGAGKEISLTVLRDGRSHVVKML-QADDSGLSSKT-EL 381
QY 407 PYTEQOQSFTSVESAGITLTQHTDSSGKHLVVVRV--SDAERAGLRHGEDEILAV 459
DB 382 PALD-----GATLKQYDAKGVKGIETITKIQPNSLAQR-GLKSGDIIIGI 425

```

```

RESULT 8
US-08-480-993-2
Sequence 2, Application US/08480993
Patent No. 5721115
GENERAL INFORMATION:
APPLICANT: Krievan, Howard C.
APPLICANT: Samuel, James E.
APPLICANT: No. 5721115berg, Nils T.
TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pernie & Edmonds

```

```

STREET: 1155 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,993
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/903,079
FILING DATE: 22-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,966
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/631,698
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7969-050
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-993-2

```

```

Query Match 27.5%; Score 646; DB 1; Length 463;
Best Local Similarity 38.9%; Pred. No. 6,6e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;

```

```

QY 54 LPDPAQLVQSEGPAVNIQAAPARTONGSNAETDS-DPLADSDP--FYEFF-KRLVFN 109
DB LPFSVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSRSPFLDDIPEEFKFFGDRFAEQ 85
QY 110 MEIIPQEAADGGGLNFGSGFII-SKNGYILTNTHVVAAGMSIKVLLNDRKREYAKLIGSD 168
DB 86 FGGRGSKRNPRGL--GSGVTIINASKGYVLTNNHVIDGADKITVQLDGRERFAKLVGKD 143
QY 169 VQSDVALLKIDATELPELVKIGNPKNLKPGEWVAAIGAPFGFDSNVTAGIVSAKGRSLPN 228
DB 144 EQSDIALVQLEKPSNLTEIKFADSDKLAVGDFVTAIGNPFGIGQVTSIGIVSALGRSTGS 203
QY 229 ES--YTFPIQTVAIINPNSGGPLFNKQGVVINSQIYSRSGFMGISFALPIDVMNV 286
DB 204 DSGTYENYIQTDAVNRNNSGALVNLNGELIGINTALISPSGNAAGIAFAIPNSQASNL 263
QY 287 AEOLKNTKXVORGOGLVITQEVSYGLAQSFGLDKASGALIAKILPGSPAERGLQAGDIY 346
DB 264 VQQLLEFGVRRGLGIRKIGELNADLAKAFNVSAQGAFFSEVLKPSAAEKAKGLAKAGDII 323
QY 347 LSLDGEIRSSGDLPMVVGATTPGKEVSLGWRKGEETIRAKLGNAAEHTGASSTDEA 406
DB 324 TANNGQKISSFAEIRAKIATTGAGKEISLTVLRDGRSHVVKML-QADDSGLSSKT-EL 381
QY 407 PYTEQOQSFTSVESAGITLTQHTDSSGKHLVVVRV--SDAERAGLRHGEDEILAV 459
DB 382 PALD-----GATLKQYDAKGVKGIETITKIQPNSLAQR-GLKSGDIIIGI 425

```

RESULT 9

US-07-903-079B-2
; Sequence 2, Application US/07903079B
; Patent No. 5843463
; GENERAL INFORMATION:
; APPLICANT: Kriivan, Howard C.
; APPLICANT: Samuel, James E.
; APPLICANT: No. 5843463berg, Nils T.
; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/903,079B
; FILING DATE: 22-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,966
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/631,698
; FILING DATE: 21-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mierock, S. Leelie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7969-004
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8664/9741
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; 7-903-079B-2

Query Match 27.5%; Score 646; DB 2; Length 463;
Best Local Similarity 38.9%; Pred. No. 6,6e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;
54 LPDPAQLVQSEGPVVNIQAAPAPRTONGSGNAETDS-DPLADSDP--FYEFF-KRLVFN 109
28 LPSFVSEQNSLAPMLRKVQ--PAVVTLSVEGKAKVDSRSPFLDDIPEEFKFFGDRFAEQ 85
110 MEIITPOEADDDGLNFGSGFIT--SKNGYILTNTHYVAVGMSIKYLLNDKREYTAKLISD 168
86 FGGRESKSNFRL--GSGVITINASKGYVLTNNHVIDGADKITVQLDGRFKAQLVKD 143
169 VQSDVALLKIDATEELPVLVIGNPKNLKPGMWVAIGAPGFDSVTAGIVSAGKRSLLPN 228
144 EOSDIALVOLERKSNLTKETKPADSKLRVDFVYALGNPGLQGTVTSGIVSALGRSTGS 203
229 ES--VTFPIQTVDVAINPGNSGGLPFLKGGVVGINSQIYSRSGGPMGISPAIPIDVAMV 286
204 DSGTYENYIQTDAAVVRGNSGALVNLNGELIGINTAIIISPSGNAGIAFAIPSNQASNL 263
287 AEQLKNTGVQNGQLGVIIQEVSYGLAQSGFLDKASGLAKLITPGSPAEAGLQAGIV 346
264 VQOILIEFGVRRGLLGIKGGELINADLAKAFNVSAQOGAFVSEVLPKSAAERAGIKAGDI 323

QY 347 LSLDGEIRSSGDLPPWVGAITPKEVSLGWRKGEIITIKAGNAAEHTGASSKTDEA 406
DB 324 TANNQKISSFAIRAKIATTAAGKEISLTFYLRGKSHDVQMKL-QADDSQLSSKT-EL 381
QY 407 PYTEQSGTFSVSAGITLQTHDSSGKLNVVAV---SDAAERAGLRHGEILAV 459
DB 382 PALD-----GATLNDYDAKGVGIEITRIQPNSLAQR-GLKSGDIIIGI 425

RESULT 10

US-08-350-741-2
; Sequence 2, Application US/08350741
; Patent No. 5804194
; GENERAL INFORMATION:
; APPLICANT: DOUGAN G.
; APPLICANT: CHARLES I.G.
; APPLICANT: HORMACHE C.E.,
; APPLICANT: JOHNSON K.S.,
; APPLICANT: CHATFIELD S.N.
; TITLE OF INVENTION: LIVE VACCINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON and VANDERHYE PC
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/350,741
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952,737
; FILING DATE: 09-MAY-1994
; APPLICATION NUMBER: US 07/952,737
; FILING DATE: 30-NOV-1992
; APPLICATION NUMBER: GB 9007194.5
; FILING DATE: 30-MAR-1990
; APPLICATION NUMBER: PCT/GB91/00484
; FILING DATE: 28-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-158
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-350-741-2

Query Match 26.9%; Score 633.5; DB 1; Length 475;
Best Local Similarity 34.1%; Pred. No. 1e-50;
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;
QY 44 TKDDGSVSMLLPPFAQLVQSEGPVV--NIQAAPAPRTONGSGNAE---TDSPLADSD 97
DB 29 TSSAMTAQOMPLALMLEKVMPSVVSINVBGTTVTPRPRPQGFPGDSPPFCQDS 88
QY 98 PVEFFPKRLVPMNPEIPOEFADDG-----LNFSGFTI--SKNGYILTNTHYVAVGK 148

Db 89 PF-----QNSPFC--QGGGNGNGGQOQKFMALGSGVILIIDAAGYVVTNNHVVNDAS 138
 QY 149 SIKVLNDRKREYTAKLIGSDVQSDVALLKIDATEELPVYKIGNPKNLKRGEMVAAGAPF 208
 Db 139 VIKVQLSDGRKFDKAVGKDPKRSIDIALIQIONPKNLTAIKLADSDALRVGDYVAIGNPF 198
 QY 209 GFDNSVTAGIVSAKGRS-LPNESYTPFIOTDVAINPNSGGPLFNLKGOVVGINSQIYSR 267
 Db 199 GLGETVTSIGIVSALGRSGNLVNNYENFIOTDAINRNGNSGALVNLNGELIGINTAILAP 258
 QY 268 SGGFMGISFAIPIDVAMNVAEQLKNTGKVGORGLGVIIEVSYGLAOSFGLDKASGALIA 327
 Db 259 DCGNIGIGFALPISNMVKNLTSQWVEYGVKRGELIGMTELSNSELAKMKVDKQAGAFVS 318
 QY 328 KILGSPAERAQLOAGDIVLSLDGGEIRSSGDLPMVVGATTPGKEVSLGVWRKGEIITIK 387
 Db 319 QVMPNSSAAKAGIKAGDVITSLNGKPISSFALPRAQVGTMPVGSKISLGLRKGKAITVN 378
 QY 388 AKLGNAEHHTGASSTIDEAPYTEQSGTSFVESA-----GITLQTHTDSGKHLVYVR 440
 Db 379 LEL-----QSSQSQVDSSTIFSGIEGAEKSNKGQDKGVVSVYK 418
 QY 441 VSDAERAGLRHGDIELAVRASP 463
 Db 419 ANSPAAQIGLKKGDVITIGANQOP 441

RESULT 11
 US-08-463-875A-2
 ; Sequence 2, Application US/08463875A
 ; Patent No. 5980907
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGAN, Gordon
 ; APPLICANT: CHARLES, Ian G.
 ; APPLICANT: HORNAECHE, Carlos E.
 ; APPLICANT: JOHNSON, Kevin S.
 ; APPLICANT: CHATFIELD, Steven N.
 ; TITLE OF INVENTION: LIVE VACCINES
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON and VANDERHYE PC
 ; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22201-4714
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/463.875A
 ; FILING DATE: 05-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/340,741
 ; FILING DATE: 07-DEC-1994
 ; APPLICATION NUMBER: US 07/952,737
 ; FILING DATE: 30-NOV-1992
 ; APPLICATION NUMBER: GB 9007194.5
 ; FILING DATE: 30-MAR-1990
 ; APPLICATION NUMBER: PCT/GB91/00484
 ; FILING DATE: 28-MAR-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: WILSON, MARY J.
 ; REGISTRATION NUMBER: 32,955
 ; REFERENCE/DOCKET NUMBER: 117-158
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 816-4100
 ; TELEFAX: (703) 816-4100
 ; TELEX: 200797 NIXN UR
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 475 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-463-875A-2

Query Match 26.9%; Score 633.5; DB 2; Length 475;
 Best Local Similarity 34.1%; Pred. No. 1e-50;
 Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;

QY 44 TXDDGSVSMLLPFPQAVQSEBPAVY--NIOAAAPRTQNGSNAE----TDSPLASD 97
 Db 29 TSSSAMTAQOMPSLAPLEKVPSEVSINVEGSTTVTPMPRRFQGFPGDSFPCOQGS 88
 QY 98 PFYEFFKRLVPMNPEIQEADQD-----LNFSGFTII-SKNGYILNTNTHVAVGNG 148
 Db 89 PF-----QNSPFC--QGGGNGNGGQOQKFMALGSGVILIIDAAGYVVTNNHVVNDAS 138
 QY 149 SIKVLNDRKREYTAKLIGSDVQSDVALLKIDATEELPVYKIGNPKNLKRGEMVAAGAPF 208
 Db 139 VIKVQLSDGRKFDKAVGKDPKRSIDIALIQIONPKNLTAIKLADSDALRVGDYVAIGNPF 198
 QY 209 GFDNSVTAGIVSAKGRS-LPNESYTPFIOTDVAINPNSGGPLFNLKGOVVGINSQIYSR 267
 Db 199 GLGETVTSIGIVSALGRSGNLVNNYENFIOTDAINRNGNSGALVNLNGELIGINTAILAP 258
 QY 268 SGGFMGISFAIPIDVAMNVAEQLKNTGKVGORGLGVIIEVSYGLAOSFGLDKASGALIA 327
 Db 259 DCGNIGIGFALPISNMVKNLTSQWVEYGVKRGELIGMTELSNSELAKMKVDKQAGAFVS 318
 QY 328 KILGSPAERAQLOAGDIVLSLDGGEIRSSGDLPMVVGATTPGKEVSLGVWRKGEIITIK 387
 Db 319 QVMPNSSAAKAGIKAGDVITSLNGKPISSFALPRAQVGTMPVGSKISLGLRKGKAITVN 378
 QY 388 AKLGNAEHHTGASSTIDEAPYTEQSGTSFVESA-----GITLQTHTDSGKHLVYVR 440
 Db 379 LEL-----QSSQSQVDSSTIFSGIEGAEKSNKGQDKGVVSVYK 418
 QY 441 VSDAERAGLRHGDIELAVRASP 463
 Db 419 ANSPAAQIGLKKGDVITIGANQOP 441

RESULT 12
 US-08-278-091-2
 ; Sequence 2, Application US/08278091
 ; Patent No. 5506139
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: OOMEN, Raymond P.
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/278,091
 ; FILING DATE: 21-JUL-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

```

? NAME: Stewart, Michael I
? REGISTRATION NUMBER: 24,973
? REFERENCE/DOCKET NUMBER: 1038-371
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (416) 595-1155
? TELEFAX: (416) 595-1163
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 463 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-08-278-091-2

```

```

Query March 26.9%; Score 632; DB 1; Length 463;
Beat Local Similarity 38.5%; Pred. No. 1,3e-50;
Matches 160; Conservative 75; Mismatches 155; Indels 28; Gaps 12

QY      54 LPDFAQLVQSEGPAAVNIQAAPAPRTONGSGNAETDS-DPLADSDP-FYEFF-KRLVFN 109
      28 LPFSFSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSSPFLDIPPEFFKFFGDRFAEQ 85
DB      110 MPETPOEADDDGGINFSGCFII-SNGGYILNTNHYVAAGMSIKVLNDRKRYTALIGSD 168
      86 FGGGESGRNRRGL--GSGVILINSKGVVLTNNHVIDADKITYQLQGRERKALVQKD 143
QY      169 VQSDVALLIKIDATELPPVKIIGNPKNLKPGEVVAIGAPFGDNSVTAGISAKRSILPN 228
      144 ELSDIALVQLEKPSNLTEIKFADSDKLAVGDPFTVALIGNPFGIGQVVTSGIYSALGRSISGS 203
QY      229 ES--YTPPIQTDVAINPNSGGPLFNLKGQVVGINSQIYSNCGCMGISPAIPIDVAMV 286
      204 DSGTYENTIQTDAAVNRNNSGALVNLNGELIGINTALIISSGGVAGAIAFAPISNQAQNL 263
QY      287 AEQLKNTKGVORGQGVIIIOEVSYGIAQSFGLDKASGALINKILPGSPAERAGLOAGDIV 346
      264 VQQLLEFGQVNRGLIGIKGELMADLAKAFVNSAQQAFVEVLPKSAEKRAGLKAGDI 323
QY      347 LSLDGEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEIITIKAGNAAEHTGASSKTDEA 406
      324 TANNQOKISSFAEIRAKIATTGAGKEISLTYLRDCKSHDVMKTL-QADDSQOLSSKT-EL 381
QY      407 PYTGQSGSTFVESAGITLQTHHTSSGHLVVVRV---SDAERAGLRHGGELLAV 459
      382 PALD-----GATLKDIYDAKGVKGIETIKIQPNSLAQR-GLKSGDIIIGI 425
DB      382 PALD-----GATLKDIYDAKGVKGIETIKIQPNSLAQR-GLKSGDIIIGI 425

SUBMIT 13
8-483-859-2
Sequence 2, Application US/08483859
Patent No. 5656436
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H..
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
TITLE OF INVENTION: Reduced Protease Activity
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```

1      APPLICATION NUMBER:  US/08/483,859
2      FILING DATE:  07-JUN-1995
3      CLASSIFICATION:  435
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER:  US 08/296,149
6      FILING DATE:  26-AUG-1994
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER:  US 08/278,091
9      FILING DATE:  21-JUL-1994
10     ATTORNEY/AGENT INFORMATION:
11     NAME:  Stewart, Michael I.
12     REGISTRATION NUMBER:  24,973
13     REFERENCE/DOCKET NUMBER:  1038-495  MIS:VG
14     TELECOMMUNICATION INFORMATION:
15     TELEPHONE:  (416) 595-1155
16     TELEFAX:  (416) 595-1163
17     INFORMATION FOR SEQ ID NO:  2:
18     SEQUENCE CHARACTERISTICS:
19     LENGTH:  463 amino acids
20     TYPE:  amino acid
21     STRANDEDNESS:  single
22     TOPOLOGY:  linear
23
24     US-08-483-859-2

```

Query Match 26.9%; Score 632; DB 1; Length 463;
Best Local Similarity 38.5%; Pred. No. 1.38-50;
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12

QY 54 LPDFAQLVQSEGPAVNNIQAAPAPRTONGSGNAETDS-DPLADSDP--FYEEF-KRLVPN 109
DB 28 LPSFVSEQNSLAPMLEKIQ--PAVVTLSEVQKAKAYDSRSPLDDIPEBFKPFEGDRAEQ 85

QY 110 MPRLPGEADGGINPSSGPII-SKNGYILTNTHVVAQMGSIKVLNDKERYTAKLIGSD 168
DB 86 FGGRGSEKRNFRGI--GSGVIYNASKGYVLTNNHVIDEADKITVQLQDGRFPAKLVEKD 143

QY 169 VQSDVALLKTIIDATEELPVVVKIGNPKNLKPGEWVAIIGAPFGFDDSVYTAGISAKGRSLPN 228
DB 144 ELSDIALVQLKESPNLTFIKFADSDKLRVGDFTVAIGNPFGCGTIVTSIGVIALGRSGS 203

QY 229 ES--YTFPIQTDVAINPQNSGGPLENLKQGVVINSQIYRSRGGFMGISAPIPIDVAMNV 286
DB 204 DSGTYENVYITQDAVNRNGNSGGALVNLNGELLIGINTAIISPSGNAGIAFAPISNQAENL 263

QY 287 AEQLKNTGKYQVQGLGVIIQEVSYGLAQSGELDKPASGALLKIIIPGSPAEPAQIAGDIY 346
DB 264 VQQLIEFGQVRRGLIGIKGGLINDAKAFNVASAQQAFFVEVLPKSAAEKAGAKAGDII 323

QY 347 LSLDGGIRRSQGLPVWVGAIITPEKEVSLGWRRGEGEITIKAKIGNAAEHNGASSTDEA 406
DB 324 TANMQKISSPAELRAKIATTGAKKEISLTLYLRGKSHDVMKQ--QADDSQSLSTKT-EL 381

QY 407 PYTEQSGSTFVESAGITLTQHTDSSGKHLVVAV--SDAAEPAGLRHGEIILAV 459
DB 382 PALD-----GATLKDYDAKGVKGIETIKIQFNSLAAQR-GLKSGIIIGI 425

RESULT 14
US-08-472-173-2
Sequence 2, Application US/08472173
Patent No. 5665353
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: COMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
NUMBER OF INVENTION: Reduced Protease Activity
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue

CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/472,173
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/296,149
 FILING DATE: 26-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/278,091
 FILING DATE: 21-JUL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael I.
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 463 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-472-173-2

Query Match 26.9%; Score 632; DB 1; Length 463;
 Best Local Similarity 38.5%; Pred. No. 1.3e-50;
 Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

54 LPDPAQLVQSEGPVAVNIQAAPARTONGSGNAETDS--DPLADSDP--FYEPF-KRLVFN 109
 28 LPSPVSEQNSLAPMLEKVQ--PAVVTLSVSGKAKVDSRSPFLDDIPEEFKFFGGRFAEQ 85
 110 MPEIPOEADDDGGLNFGSGFII--SKNGYILTNTHVAVGMGSIKYLLNDKREYAKLIGSD 168
 86 FGGRGESKRNFRGL--GSGVITINASKGYVLTNNHVIDEADKITVQLQDGRERFAKLVGKD 143
 169 VQSDVALLKIDATEELPVVKIGNPKNLKPGEWVAALGAPFGFDNSVTAGIVSAKGRSLPN 228
 144 ELSDIALVQLEKPSNLTETIKFADSDKLRYGDFVTAIGNPFGAQVTSIGIVSALGRSTGS 203
 229 ES--YTPRIQTDVAINPNSGGPFLNFKGOVVGINSQIYSRSGGFMGISFAIPIDVAMV 286
 204 DSGTYENYIOTDAAVNRGNSGALVNLNGLIGINTAIISSPGNAGIAFAIPSNQASNL 263
 287 AEQLKNTGKVOQGOLGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346
 264 VQQLIERGQVRRLGLIGIKGELNADLAKAFNVSAQGAFAVEVLPKSAEKAAGLKAQDII 323
 347 LSLDGEIRSSGDLPMVWGAIITPGKEVSLGVWRKGEETITAKLGNAAEHTGASKTDEA 406
 324 TAMNGOKISSFAIRAKIATGTGAKESILTYLRDGSMDVKKML-QADSSQLSKST-EL 381
 407 PYTEQSGTFSVSAGITLQHTDSSGKHLVVVAV--SDAARAGLRHGDILLAV 459
 382 PALD-----GATLKDYDARKVGKIEITKIQPNLSIAQR-GLKSGDIIIGI 425

RESULT 15
 US-08-487-167-2
 ; Sequence 2, Application US/08487167
 ; Patent No. 5863302
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M.

APPLICANT: YANG, Yan-Ping
 APPLICANT: CHONG, Pele
 APPLICANT: COMEN, Raymond P.
 APPLICANT: KLEIN, Michel H.
 TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
 TITLE OF INVENTION: Reduced Protease Activity
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sim & Mcburney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,167
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/296,149
 FILING DATE: 26-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/278,091
 FILING DATE: 21-JUL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael I.
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 463 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-487-167-2

Query Match 26.9%; Score 632; DB 2; Length 463;
 Best Local Similarity 38.5%; Pred. No. 1.3e-50;
 Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

54 LPDPAQLVQSEGPVAVNIQAAPARTONGSGNAETDS--DPLADSDP--FYEPF-KRLVFN 109
 28 LPSPVSEQNSLAPMLEKVQ--PAVVTLSVSGKAKVDSRSPFLDDIPEEFKFFGGRFAEQ 85
 110 MPEIPOEADDDGGLNFGSGFII--SKNGYILTNTHVAVGMGSIKYLLNDKREYAKLIGSD 168
 86 FGGRGESKRNFRGL--GSGVITINASKGYVLTNNHVIDEADKITVQLQDGRERFAKLVGKD 143
 169 VQSDVALLKIDATEELPVVKIGNPKNLKPGEWVAALGAPFGFDNSVTAGIVSAKGRSLPN 228
 144 ELSDIALVQLEKPSNLTETIKFADSDKLRYGDFVTAIGNPFGAQVTSIGIVSALGRSTGS 203
 229 ES--YTPRIQTDVAINPNSGGPFLNFKGOVVGINSQIYSRSGGFMGISFAIPIDVAMV 286
 204 DSGTYENYIOTDAAVNRGNSGALVNLNGLIGINTAIISSPGNAGIAFAIPSNQASNL 263
 287 AEQLKNTGKVOQGOLGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346
 264 VQQLIERGQVRRLGLIGIKGELNADLAKAFNVSAQGAFAVEVLPKSAEKAAGLKAQDII 323
 347 LSLDGEIRSSGDLPMVWGAIITPGKEVSLGVWRKGEETITAKLGNAAEHTGASKTDEA 406
 324 TAMNGOKISSFAIRAKIATGTGAKESILTYLRDGSMDVKKML-QADSSQLSKST-EL 381
 407 PYTEQSGTFSVSAGITLQHTDSSGKHLVVVAV--SDAARAGLRHGDILLAV 459

Db 382 PAJD-----GATLKDYDAKGVKGIEITRKIQPNSLAAQR-GLKSGDIIIGI 425

Search completed: July 11, 2003, 10:34:33
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:33:18 ; Search time 54 Seconds
(without alignments)
1002.504 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKKYQYFALALCALIAG.....ERAGLRHDELLAVRASPQ 465

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 segs, 116419773 residues 445758
1 number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2234.5	95.0	498	10 US-09-388-089B-11	Sequence 11, Appl
2	2130.5	90.5	475	10 US-09-388-089B-12	Sequence 12, Appl
3	1992.5	84.7	447	10 US-09-388-089B-2	Sequence 2, Appl
4	980.5	42.1	460	9 US-09-975-719-132	Sequence 132, App
5	731.5	31.1	503	10 US-09-752-385-8	Sequence 8, Appl
6	705	30.0	474	9 US-10-210-296-7	Sequence 7, Appl
7	559.5	23.8	387	9 US-09-895-913A-120	Sequence 120, App
8	476.5	20.3	619	9 US-10-156-761-10652	Sequence 10652, A
9	455.5	19.4	472	9 US-10-156-761-11757	Sequence 11757, A
10	453	19.3	549	9 US-09-712-363-190	Sequence 190, App
11	443.5	18.8	403	9 US-10-102-806-552	Sequence 552, App
12	442.5	18.8	413	9 US-09-738-626-4478	Sequence 4478, App
13	418	17.8	464	9 US-09-712-363-182	Sequence 182, App
14	414.5	17.6	458	9 US-10-197-634-1	Sequence 1, Appl
15	414	17.6	348	9 US-09-796-753-34	Sequence 34, Appl
16	412.5	17.5	453	9 US-09-796-753-32	Sequence 32, Appl
17	412.5	17.0	476	10 US-09-935-390A-37	Sequence 37, Appl
18	399.5	15.2	286	10 US-09-764-898-256	Sequence 256, App
19	356.5	15.2	411	9 US-10-156-761-11579	Sequence 11579, A

20	348	14.8	330	10 US-09-764-898-184	Sequence 184, App
21	313.5	13.3	178	9 US-09-969-384-18	Sequence 18, Appl
22	312	13.3	9	US-09-712-363-161	Sequence 161, App
23	285.5	12.1	255	9 US-09-866-050A-694	Sequence 694, App
24	242	10.3	51	10 US-09-388-089B-16	Sequence 16, Appl
25	213.5	9.1	596	10 US-09-287-849-26	Sequence 26, Appl
26	213.5	9.1	729	10 US-09-287-849-2	Sequence 2, Appl
27	206	8.8	361	9 US-10-197-634-15	Sequence 15, Appl
28	184	7.8	397	9 US-09-712-363-280	Sequence 12342, A
29	156	6.6	394	9 US-10-156-761-13542	Sequence 12342, A
30	141	6.0	399	9 US-10-156-761-12129	Sequence 12129, A
31	137.5	5.8	767	10 US-09-919-497-59	Sequence 59, Appl
32	136.5	5.8	452	9 US-10-156-761-15025	Sequence 15025, A
33	136	5.8	2037	9 US-09-951-402-3	Sequence 3, Appl
34	136	5.8	2037	10 US-09-951-402-3	Sequence 3, Appl
35	136	5.8	2037	10 US-09-922-101-3	Sequence 3, Appl
36	131	5.6	68	10 US-09-864-761-34281	Sequence 34281, A
37	130	5.5	724	9 US-10-211-962-21	Sequence 21, Appl
38	129.5	5.5	397	9 US-09-738-626-3831	Sequence 3831, Ap
39	128.5	5.5	204	10 US-09-925-300-1397	Sequence 1397, Ap
40	128	5.4	267	10 US-09-735-705-352	Sequence 352, App
41	128	5.4	267	10 US-09-850-716A-352	Sequence 352, App
42	128	5.4	267	10 US-09-897-778-352	Sequence 352, App
43	128	5.4	683	10 US-09-841-132-357	Sequence 357, App
44	126.5	5.4	434	9 US-10-156-761-10099	Sequence 10099, A
45	125	5.3	304	9 US-10-012-896-835	Sequence 835, App

ALIGNMENTS

RESULT 1					
US-09-388-089B-11					
; Sequence 11, Application US/09388089B					
; Patent No. US2002018782A1					
; GENERAL INFORMATION:					
; APPLICANT: Jackson, W.					
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID					
; TITLE OR INVENTION: SEQUENCE AND USES THEREOF					
; FILE REFERENCE: 7969-083					
; CURRENT APPLICATION NUMBER: US/09/388, 089B					
; CURRENT FILING DATE: 1999-08-31					
; NUMBER OF SEQ ID NOS: 20					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 11					
; LENGTH: 498					
; TYPE: PRT					
; ORGANISM: Neisseria meningitidis					
US-09-388-089B-11					
Query Match					
Best Local Similarity 95.0%; Score 2234.5; DB 10; Length 498;					
Matches 444; Conservative 5; Mismatches 13; Indels 1; Gaps 1;					
Qy	1	VFKKYQYFALALCALIAGCEKAGSFRGADKKASVETIEHTKDDGSVMLPDPFQOL	60		
Db	1	VFKKYQYFALALCALIAGCEKAGSFRGADKKASVETIEHTKDDGSVMLPDPFQOL	60		
Qy	61	VOSEGPVAVNIQAAPARTONGSGNAETDSDPLADSPFEFFFRRLVPMMEIPOEEADD	120		
Db	61	VOSEGPVAVNIQAAPARTONGSGNAETDSDPLADSPFEFFFRRLVPMMEIPOEEADD	120		
Qy	121	GGINFGSGFTISKNGYILNTHTVAVAGSITKVLINDREBYTAKLIGSDVQSDVALLKIDA	180		
Db	121	GGINFGSGFTISKNGYILNTHTVAVAGSITKVLINDREBYTAKLIGSDVQSDVALLKIDA	180		
Qy	181	TEELPVVKIGNPKNLKRGVVAALIGAPGSDNSTAGIVAKGSLNVESTTPFQDTVA	240		
Db	181	TEELPVVKIGNPKNLKRGVVAALIGAPGSDNSTAGIVAKGSLNVESTTPFQDTVA	240		
Qy	241	INFGNSGPFLENKGVGVINSQIYSRSGFMGIFALPIDVANVAEQLKNTGKGVQGO	300		
Db	241	INFGNSGPFLENKGVGVINSQIYSRSGFMGIFALPIDVANVAEQLKNTGKGVQGO	300		

Db 240 INFGSGGLFNLKQGVVINSQIYRSRSGGFMGISPAIPIDVANNVAEQLKNTGKVORQ 299
Qy 301 LGVITIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGIQAQDIYLSLDGGEIRSSGDL 360
Db 300 LGVITIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGIQAQDIYLSLDGGEIRSSGDL 359
Qy 361 PVMVGAITPGEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQSGTFVSES 420
Db 360 PVMVGAITPGEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQSGTFVSES 419
Qy 421 AGTTLQHTDSSGKHLVVRVSDAERAGLRGDEILAVRASP 463
Db 420 AGTTLQHTDSSGKHLVVRVSDAERAGLRGDEILAVQVVP 462

RESULT 2

US-09-388-089B-12
; Sequence 12, Application US/09388089B
; Patent No. US20020018782A1

GENERAL INFORMATION:

APPLICANT: Jackson, W.
APPLICANT: Harris, A.
TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID
FILE REFERENCE: 7969-083
CURRENT APPLICATION NUMBER: US/09/388, 089B
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 12
LENGTH: 475
TYPE: PRT
ORGANISM: Neisseria meningitidis
US-09-388-089B-12

Query Match 90.5%; Score 2130.5; DB 10; Length 475;
Best Local Similarity 96.4%; Pred. No. 2,1e-163;
Matches 424; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

Qy 24 ACSPFGADKKEASFVRIEHTKDDGSVMCLPFPALQVSEGPVAVNIOAAPARTONGS 83
Db 1 ACSPFGADKKEASFVRIEHTKDDGSVMCLPFPALQVSEGPVAVNIOAAPARTONGS 60
Qy 84 GNAETDSDPLADSDPFYEFKRLVPMNPBIPQEBADGGINFGSGFIISXNGYILTNTHV 143
Db 61 SNAETDSDPLADSDPFYEFKRLVPMNPBIPQEBADGGINFGSGFIISXNGYILTNTHV 120
Qy 144 VAGMGSIKYLNDKREYTAQLIGSDVQSDVALLKIDATEELPVVKIGNPKLKEGEVAA 203
Db 121 VAGMGSIKYLNDKREYTAQLIGSDVQSDVALLKIDATEELPVVKIGNPKLKEGEVAA 180
Qy 204 IGAPFGDNSVTAGIYSAKGRSLPNESYTPFIOTDVAINFGSGGLFNLKQGVVINSQ 263
Db 181 IGAPFGDNSVTAGIYSAKGRSLPNESYTPFIOTDVAINFGSGGLFNLKQGVVINSQ 239
Qy 264 IYSRSGGFMGISPAIPIDVANNVAEQLKNTGKVORQGLVITIEVSYGLAOSFGLDKASG 323
Db 240 IYSRSGGFMGISPAIPIDVANNVAEQLKNTGKVORQGLVITIEVSYGLAOSFGLDKASG 299
Qy 324 ALIAKILPSSPARAGIQAQDIYLSLDGGEIRSSGDLPMVVGAIITPGEKESVSLGVMRKGE 383
Db 300 ALIAKILPSSPARAGIQAQDIYLSLDGGEIRSSGDLPMVVGAIITPGEKESVSLGVMRKGE 359
Qy 384 ITIKKILGNAAEHTGASKTDEAPYTEQSGTFVSESAGITLQHTDSSGKHLVVRVSD 443
Db 360 ITIKKILGNAAEHTGASKTDEAPYTEQSGTFVSESAGITLQHTDSSGKHLVVRVSD 419
Qy 444 AERAGLRGDEILAVRASP 463
Db 420 AERAGLRGDEILAVQVVP 439

RESULT 3

US-09-388-089B-2
; Sequence 2, Application US/09388089B
; Patent No. US20020018782A1
; GENERAL INFORMATION:
APPLICANT: Jackson, W.
APPLICANT: Harris, A.
TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID
FILE REFERENCE: 7969-083
CURRENT APPLICATION NUMBER: US/09/388, 089B
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 447
TYPE: PRT
ORGANISM: Neisseria spp.
US-09-388-089B-2

Query Match 84.7%; Score 1992.5; DB 10; Length 447;
Best Local Similarity 96.4%; Pred. No. 2,5e-152;
Matches 397; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

Qy 52 MLDPFQVLVSEGPVAVNIOAAPARTONGSGNAETSDPLADSDPFYEFKRLVPMNP 111
Db 1 MLDPFQVLVSEGPVAVNIOAAPARTONGSGNAETSDPLADSDPFYEFKRLVPMNP 60
Qy 112 EIPQEBADGGINFGSGFIISXNGYILTNTHVAVAGMSIKYLNDKREYTAQLIGSDVOS 171
Db 61 EIPQEBADGGINFGSGFIISXNGYILTNTHVAVAGMSIKYLNDKREYTAQLIGSDVOS 120
Qy 172 DVALLKIDATEELPVVKIGNPKLKEGEVAAIGAPFGDNSVTAGIYSAKGRSLPNESY 231
Db 121 DVALLKIDATEELPVVKIGNPKLKEGEVAAIGAPFGDNSVTAGIYSAKGRSLPNESY 179
Qy 232 TPIQTDVAINFGSGGLFNLKQGVVINSQIYRSRSGGFMGISPAIPIDVANNVAEQLK 291
Db 180 TPIQTDVAINFGSGGLFNLKQGVVINSQIYRSRSGGFMGISPAIPIDVANNVAEQLK 239
Qy 292 NTGKVVORQGLVITIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGIQAQDIYLSLDG 351
Db 240 NTGKVVORQGLVITIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGIQAQDIYLSLDG 299
Qy 352 GEIRSSGDLPMVVGAIITPGEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQ 411
Db 300 GEIRSSGDLPMVVGAIITPGEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQ 359
Qy 412 QSGTFVSESAGITLQHTDSSGKHLVVRVSDAERAGLRGDEILAVRASP 463
Db 360 QSGTFVSESAGITLQHTDSSGKHLVVRVSDAERAGLRGDEILAVQVVP 411

RESULT 4

US-09-975-719-132
; Sequence 132, Application US/09975719
; Publication No. US20030022349A1

GENERAL INFORMATION:

APPLICANT: Ausubel, Frederick M.
APPLICANT: Rahme, Laurence G.
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
FILE REFERENCE: 00786/361003
CURRENT APPLICATION NUMBER: US/09/975, 719
CURRENT FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 09/199,637
PRIOR FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: US 60/066,517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 132
LENGTH: 460
TYPE: PRT

ORGANISM: Pseudomonas aeruginosa
US-09-975-719-132

Query Match 42.1%; Score 990.5; DB 9; Length 460;

Best Local Similarity 52.5%; Pred. No. 1.3e-71; Indels 33; Gaps 11;

Matches 220; Conservative 55; Mismatches 111;

Db 54 LPDPAQVQSEGPAAVNI---QAAPARTONGSNAETSDPLADDP-FYEFKRLVPMPI 109

27 LPDFTPLVQSPAVVNISTRQKLP---DRAMARQQLSPDLEGLPMPMRDLERTIPQ 82

110 MPEIR---QEEADDDGLNFGSGFFISKNGYILTNTHVAVAGMSIKYLLNDKREYTAALIG 166

83 VPRNRGQGRQAQ---SLGSGFFIISNDGYILTNHVVADADAILVRLSRSHKAKLVG 138

167 SDVQSDVALLKIDATEELPVVKGKPNKPKPGEWAAIAGAPFGFDSVYTAGIYSAGRS 226

139 ADPRSDVAVLKEA-KNPLTKLGDNNKLVGEMVVAIAGSPFGFDSVYTAGIYSAGRS 197

227 PNESTYPTIQTDAINPGNSGGPLFNLKGQVAGINSQIYSRSGFGMGSFAIPDIVAMNY 286

198 PNESTYPTIQTDAINPGNSGGPLNLBEGVAGINSQIYTRSGFGMGSFAIPDIVAMNY 257

287 AEOLKNTGVORGOIGVITIQEVSYGIAQSGFGLDKASGALLAKILPGSPAPRAGLQAGDIV 346

258 ADOLKAKGVSRGVLGVITIQEVSYGIAQSGFGLDKASGALLAKILPGSPAPRAGLQAGDIV 317

347 LSLDGEIRSSGDLPMVNGAITRGEVSLGWKRGKEITIKALGNAEHTTGASKTDE- 405

318 LSLNGQSI NESADLPHLVGNMKGDKINLDIVIRNGRKSLSMAVGNLDP-----DDEE 370

406 ----APYEQSGTFVSAGITLQTH--TDSSGKHLVVVRVSDA-AERAGLRHGEI 456

371 IASMAFPAERSNNLGVYVADLTAEGRKSLDYG--GVYIKYQDGPAAVYIGLRPGDIV 428

Db 371 IASMAFPAERSNNLGVYVADLTAEGRKSLDYG--GVYIKYQDGPAAVYIGLRPGDIV 428

RESULT 5
US-09-752-385-8

Sequence 8, Application US/09752385

Patent No. US20020012919A1

GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.

Regnery, Russell L.

TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae

and Methods and Compositions for Diagnosing

Rochalimaea

Henselae and Rochalimaea Quintana Infection

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.

STREET: 127 Peachtree Street, N.E., Suite 1200

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/752,385

FILING DATE: 29-Dec-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/525,310

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Soratt, Gwendolyn D.

REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 1414.624

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 688-0770

TELEFAX: (404) 688-9880

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 503 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-752-385-8

Query Match 31.1%; Score 731.5; DB 10; Length 503;

Best Local Similarity 39.5%; Pred. No. 1e-50;

Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

Db 8 FALAALCAALLAGEKAGSFFGADKKEASFVERIEHTDDGSVSLPLDPAQVQSEGA 67

15 FSALELTALEFFSGC---GSSLMTTKAHANSV-----FSSLMOQGGFADIVGQVKA 62

68 VVNIQAAPARTON---GSGNAETSDPLADSDP---FYEFKRLVPMPEI PQEADDD 120

63 VVSVQVAKSNKKKKKEMFSDPSTGPFQQLPDQHFLKRFQDFVNRDKFSNKL-QRSHRL 121

121 GGLNFGSGFFIISKNGYILTNTHVAVAGMSIYVLLNDKREYTAALIGSVQSDVALLKIDA 180

122 RPIAFGSGFFIISNDGYILTNHVVADADAILVRLSRSHKAKLVG 181

181 TEBLPVVKIGNPKLKPGEWAAIAGAPFGFDSVYTAGIYSAGRS 240

182 KRKFSYVDFGDSKLRAGDVVAIAGNPFGIAGTVAGIYSAGRGITGVVDFQIDDA 241

241 INPGNSGGPLFNLKGQVAGINSQIYSRSGFGMGSFAIPDIVAMNVAEOLKNTGVORGO 300

242 VNRGNSGGPFDNLGKVGAVNTAIFSPSGGVAGIAFAIPATANEVQQLIEKGLVGRW 301

301 LGVITQEVSYGLAQSFGGLDKASGALLAKILPGSPAPRAGLQAGDIVLSLDCGEIRSSGDL 360

302 LGVITQEVSYGLAQSFGGLDKASGALLAKILPGSPAPRAGLQAGDIVLSLDCGEIRSSGDL 360

361 PVMWGAITPGKEVSLGWKRGKEITIKALGNAEHTTGASKTDEAPT-EEQSGTFVSVE 419

361 AKRIANNSPGSTVYLVGWSKGENIKVLDSPMD--ENMKDSKYSNENGSDETLE 417

420 SAGITLQTHDSSGKHLVVVRV---SDAERAGLRHGEIILAV 459

418 DYGLIAPSDDGLG--LVITDVDPDSDADK-GIRPGDIVTV 457

Db 418 DYGLIAPSDDGLG--LVITDVDPDSDADK-GIRPGDIVTV 457

RESULT 6

US-10-210-296-7

Sequence 7, Application US/10210296

Publication No. US20030021802A1

GENERAL INFORMATION:

APPLICANT: Pfizer Products Inc.

TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED

TITLE OF INVENTION: METHODS

FILE REFERENCE: PC10589A

CURRENT APPLICATION NUMBER: US/10/210,296

CURRENT FILING DATE: 2002-08-01

PRIOR APPLICATION NUMBER: US/09/689,065

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 102

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 7

LENGTH: 474

TYPE: PRT

ORGANISM: Lawsonia intracellularis

US-10-210-296-7

Query Match 30.0%; Score 705; DB 9; Length 474;

Best Local Similarity 39.3%; Pred. No. 1.3e-48;

Matches 169; Conservative 67; Mismatches 152; Indels 42; Gaps 11;

Qy	54	PDFFQOLVQSEBPAAVNIQA-APAPRTQNGSGNAETDSDPLADSDPYE-FPRKLVPMNP	1111
Dd	28	LPNFVPLVKDASKAAVYNIISTEKKIPR-----GRTPEMEMPRGLPROFEAFEEQFEFKCP	82
Qy	112	--EIPQEBADQGLNFGSGFIISKNGYILTNTHVAVGMSIKVLL-----NDKREYTAKL	1655
Dd	83	DSQIHKKR-----SLGTGFIISSDPYIYITNNHVIGAGDSVRNLEGTSGKESLPRAEVI	1366
Qy	166	GSDVQSDVALLKIDATTEELPVVKIGIPKXULKPEEMWAAIICAPRCPFQNSYTAGIVSAAGRS	2255
Dd	137	GRDESTDALLKVKSKDSLPLYILFGNSDMEGEWELAIAGNPGILGHYVAGIILSAGRD	1966
Qy	226	LPNESYTFEPIQTDVAINPFGNSGCPULFNLGQVVINGVINSQIYSRSGGFMGISFAIPIDVAMN	2855
Dd	197	IHAGFBNDFLOTDASINPENSNGPFLNMSGQVVYGINITALMA-SG--QGIGFALPSSMAHR	2555
Qy	286	VAEQUKATGKVRQGLGVITQEVSYGCLAQSGFGLDKASGALLIAKILPGSPAPRAGLOAGDI	3455
Dd	254	IIEQLTKTKKYSRGMIGVYIQVDYNTAAALGLSQAKGALVSGVSPGDPADKGLKVGDI	3133
Qy	346	VLSLDGGEIRSSGDLPVWVGALITPCKEVEVLGWVRKCEBITIKAKLNAALNHTGASSKTE	4055
Dd	314	VTQADGKQIDASSLLKAIATATPRPSVVLKAKWRDKSKDQISTTLBERKTTSSQKQSSPE	3737
Qy	406	A-----PYTEQOQSTFQSVESAGITLQTHTDSGKRLVVVRVSDAERAAGLRHGE	4555
Dd	374	SLPGALGSLVRPLTQESKSFQVK-LGI-----GLLVEVEPNKPRSEAGIREQDI	4233
Qy	456	ILAVYASPRQ	465
Dd	424	ILSANLKFPQ	433

```

RESULT 7
US-09-895-913A-120
/ Sequence 120, Application US/09895913A
/ Patent No. US20020160456A1
/ GENERAL INFORMATION:
/ APPLICANT: Kleanthous, Harold
/ APPLICANT: Al-Garawi, Amal
/ APPLICANT: Miller, Charles
/ APPLICANT: Tomb, Jean Francois
/ APPLICANT: Oomen, Raymond P.
/ TITLE OF INVENTION: Identification of Polynucleotides
/ TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
/ TITLE OF INVENTION: Genome
/ FILE REFERENCE: 06132/043002
/ CURRENT APPLICATION NUMBER: US/09/895,913A
/ CURRENT FILING DATE: 2001-06-29
/ PRIOR APPLICATION NUMBER: US 08/881,227
/ PRIOR FILING DATE: 1997-06-24
/ NUMBER OF SEQ ID NOS: 368
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 120
/ LENGTH: 387
/ TYPE: PRF
/ ORGANISM: Helicobacter pylori
/ US-09-895-913A-120

```

	Query Match	23.8%	Score 559.5;	DB 9;	Length 387;
	Best Local Similarity	36.9%	Pred. No. 5e-37;		
	Matches 131; Conservative	73;	Mismatches 134;	Indels 17;	Gaps 7;
Oy	113 I P O E A D D G G L N F G S G F T I S K N G V I L T T H V A G M G S I K V L L - N D K R E Y T A L I G S D V Q S	171			
	: : : : : : : :				
Db	2 I P K R M E R A - - L G S G V I I S K D G Y I V T N H N V I D G A D K I K V T I P G S K E Y S A T L V G T D S B S	58			
Oy	172 D V A L L K I D A T E E L P V W K I G N P K U L K P G E M V A I G A P F G D N S V T A G I Y S A K R S - L P N E S	230			
	: : : : : : : : : : : : : :				
Db	59 D L A V I R I - T K N L L P T I K F S D S N D I S V G L V F A I G N P F G S E S V T T G A I S A L N K S G I G I N S	117			
Oy	231 Y T P I O T D V A I N P G N S G G P L F N L K G Q V V I N S Q I S R S G G E F N G I S F A I P I D V A N N V A E D L	290			
	: : : : : : : : :				

[illegible]

```

RESULT 8
US-10-156-761-10652
; Sequence 10652, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, WASHIIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10652
; LENGTH: 619
; TYPE: PRt
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10652

```

	Query Match	20.3%;	Score 476.5;	DB 9;	Length 619;	
	Best Local Similarity	32.0%;	Pred. No. 4.7e-30;			
	Matches 133;	Conservative	64;	Mismatches 142;	Indels 77;	Gaps 10;
Qy	4	KYQYFALAAALCAALLACGKAGSFFGADKKEASFVERIEHTKDDGSVMLLPDAFQVLQS	63			
Db	254	RVRLVLTGAVLVALVLSG-----GIGAVGVYLLER-----TGVTVDVLELPQAQG--ES	298			
Qy	64	EGPAAVNI-----QAAPAPRTQNSGNAETDSDPLASDPYEFEGFKLVPMMPPEIPOEBA	118			
Db	299	EERAAADSVAIGIAAALPSVTLTHVKGSAABCT-----	330			
Qy	119	DDGGLINSGEFIISKNGYIILTNTHVAVAGMS---IKYLLNDKREYTAKLIGSDVQSDVAL	175			
Db	331	-----GTGVLDLGRHILLTNHNVVBPAGSSGEISVTFSGGEPAKATVVGGRDSGYDLAV	383			
Qy	176	LKIDATEELPVVKI GNPQNLKPGEMVAALGAPFQFDNSVTAGIYSAKRSL-----	226			
Db	384	VKQSGVGGLKPMPLGNSDNOVGDPVVAIGAPFLAATVTVISGIIISAKEPITTAGGCEKGDG	443			
Qy	227	PNESYETPIOTDVAVINPGNSGGPLFNLKGVAVGINSOIYS-----RSG--GFMGISPAI	278			
Db	444	SDVSIVNALQTDAPINFGNSGGPLFDBSKARVYVINSAIRADSSSDSGQAGSITGLGPAI	503			
Qy	279	PIDVAMNVAEQLKNTGKTVQSGQLGVIIQEVSYGLAQSFGFLDKAGALIAKLIPSSPAERA	338			
Db	504	PVNAQKRVAEELINTGRATHPVIGVTLTDMDYTGDAARGVTKSNQD--SPVTRGSGCPBRA	561			
Qy	339	GLQAGDVLVLSDGEIRSSGDLPLMVQALITPQKEVSGCWKRKEEITIKALGNA	394			
Db	562	GIQAGDVITVEDGQRHSIGELLVYKIAHPRGDALLTVEEDGKEKPVTLVLSGAS	617			

```
RESULT 9
US-10-156-761-11757
; Sequence 11757, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATOKI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11757
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11757

Query Match          19.4%; Score 455.5; DB 9; Length 472;
Best Local Similarity 32.5%; Pred. No. 1.6e-28;
Matches 138; Conservative 53; Mismatches 140; Indels 93; Gaps 14;

QY 11 AALCAALLAGCEKASFFGADKKEASFVERIETHTKDDSGVSMILPDPFAOLVSEGPAYVN 70
DB 108 AILVPAALVAGVGVGIGVTLAK-----DNOSSSGSTTVSASDSGSGVGRDGTAG 158
QY 71 IQAAPART-----QNGSQAETSDPLADSDPFYEFKRLVPMPEIPOEBADGGLNF 125
DB 159 VAKALPSTVTIENAOSSSG-----EGGT-- 181
QY 126 GSGFIISKNGGILNTHYVAGM-----GSIKVLNDRREYAKLIGSDVSDVALKI-DAT 181
DB 182 GTFGEVFDKGGHIVNNHVAEVDGKLTATFPDGKNAEVAAGYDAVAVLTKAP 241
QY 182 EELPVVAKIGNPKNLKPEWVAIIGAPFGDNSVTAGIVSAK-----GRLPMESTYP 233
DB 242 SDLOPLTLDGSDKVAVDSTTALIGAPGLSTVTITGIIISAKRPVASSDGGSSSKASTMS 301
QY 234 FIOTDVAINPNSGGPLFNLKGQVVGINSQIYSRS-----GFGMISPAIPIDVAM 284
DB 302 ALQTDASINPNSGGPLLDAGSVIGINSALQSSSSGGIGSGGSGSGLGFPALPIINQAK 361
QY 285 NVAEOLKNTKQVORGQGLVIOEVSYGLAQSGFLDK-----ASGALLAKILPGSPAERAG 339
DB 362 YVAOELKTKPKVPYPVIG-----ASVLEEGTGAKKTEOGASGS--DAITPNGPAARAG 414
QY 340 LQAGDIYLSLDGEIRSSGDLPMVWGAIT--TPGEVSLGVWRKGEETITIAKLGNAEH 396
DB 415 LKRGDVTITKDDMWIDSG---PTLIGELWTRPAPATVATLTTRDOKANTTDTVTLG---ER 468
QY 397 TGAS 400
DB 469 VGDG 472

RESULT 10
US-09-712-363-190
; Sequence 190, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: EISENBERG, DAVID
; APPLICANT: ROTSTEIN, SERGIO H.
```

```
; APPLICANT: MARCOTTE, EDWARD M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-190

Query Match          19.3%; Score 453; DB 9; Length 549;
Best Local Similarity 30.3%; Pred. No. 3.1e-28;
Matches 122; Conservative 68; Mismatches 142; Indels 70; Gaps 10;
```

```
QY 4 KYQYPAALCA-ALLAGEKASFFGADKKEASFVERIETHTKDDSGVSMILPDPFAOLVQ 62
DB 196 KVSYLALGITVALVALVIG--GIGGVIG--RKTAEVVDAPFTSK----- 234
QY 63 SEGPAVNIQAAAPARTQNGSQAETSDPLADSDPFYEFK--RLVPMPEIPOEBAD 120
DB 235 -----VTLSTTGNAG-----EPAGRETKVAANAADSVTTIESVDQ 270
QY 121 GGLNFGSGFIISKNGGILNTHYVAGM-----IKVLNDRREYAKLIGSDVSDVA 174
DB 271 EGMQ-GSGVITVDGRGIVNNHVISPAANNPSQFTTVFNDKEVPANLVGRDPKTDLA 329
QY 175 LKIDATEELPVVAKIGNPKNLKPEWVAIIGAPFGDNSVTAGIVSAKGRSLP----- 227
DB 330 VLKVNVDMLTVARLGDSSKRVAGDEVLAVGAPLGRSTVTOGIVSALHHPVPLSGGSD 389
QY 228 NESYTPFIOTDVAINPNSGGPLFNLKGQVVGINSQIYSRSGGFGMISPAIPIDVAMVA 287
DB 390 TDTVIDALQTDASINHGNSGGPLIMDAQVIGINTAGKSLSDSASGLGFAIPVNEKLVVA 449
QY 288 EGLKNTGKQVORGQGLVIOEVSYGLAQSGFLDKASGALLAKILPGSPAERAGIQAQIVL 347
DB 450 NSLIDGKIHPFTLIGSTRSVSNAL-----ASQAQVANVAGSPAOKGILENDVIV 501
QY 348 SLUDGEIRSSGDLPMVWGAITPGEVSLGVWRKGEETITIAK 389
DB 502 KVGNAVADSDDEFVAVVAGQLAIGQDAPIVVAEGRHVTLTVK 543

RESULT 11
US-10-102-806-552
; Sequence 552, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS AND ANTIBODIES
; FILE REFERENCE: PA103PIC1
```

```
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 552
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-102-806-552

Query Match      18.8%; Score 443.5; DB 9; Length 405;
Beet Local Similarity 34.6%; Pred. No. 1.2e-27;
Matches 110; Conservative 55; Mismatches 96; Indels 57; Gaps 7;

58 ADLVQSEGPVAVNIQAAPAPRTQNGSGNAETDSDPLADSDPFYEFKRLVPMPEIPOE 117
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 ADVVEKIAAAVNH-----ELFRKLPSKREVP--- 136

118 ADDGGINFSGFTISKNGYILTNTHVAVAGMSIVYLNDKREYTKLIGSDVQSVALLK 177
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 -----VASGSGFTIVSEDLIVTNHVVTNKHRYKVELKNGATYEAKIKVDDEKADIALIK 191

178 IDATELPVYKIGNPKULKKEGWVAAIAGAPGFDSVTAGIVSAK---GRSLP-NESTYP 233
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 IDHGLPVLPLLRSSSELRFGEFVVALGSPSLNTVTYTGIVSTORGKELGLRNSDMD 251

234 FIOTDVAINPQNSGGLFNLKQGVVINSQIYSRSGFGMISFAIPID-----VANVAE 288
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 YIOTDAIINVNGSGPLVNDGEVIGINTLKYT-----AGISFAIPSKIKKFLTESHDR 306

289 QLKNTGKVQRGQLGVIIQEVSYGLAQSFG-----LDKASGALLAKILPGSPAERAGLQA 342
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 QAKGKAITKKKYIGIRHMSLITSSKAKELKDRHRDPDIVISGAVIIEVAPDPAEAGLKE 366

343 GDIVLSLDGEIRSSGDL 360
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
367 NDVIISINGQSVVANDV 384

RESULT 12
; US-09-738-626-4478
; Sequence 4478, Application US/09738626
; Application No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 4478
```

```
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-4478

Query Match      18.8%; Score 442.5; DB 9; Length 413;
Beet Local Similarity 32.9%; Pred. No. 1.5e-27;
Matches 129; Conservative 57; Mismatches 137; Indels 69; Gaps 10;

3 KKQYFALALCALAGCERAGSFFGADKKEASFVERIEHTKDDGSVSMLLPDAQLVQ 62
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 KEKKVIGITALLMLVGSITATGVSVG-----AAVQIG 89

63 SEGPVAVNIQAAPA-PTQNGS--GNAETDSDPLADSDPFYEFKRLVPMPEIPOEADD 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 SDSSTPVNALQEPSVQRTTAEPGSAEQVANAALVPS-----VVSIAQITRSABE 139

121 GGLNFGSGFTISKNGYILTNTHVAVAGM---GSIKVLNDKREYTKLIGSDVQSVALLK 177
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 -----GSGSIISDGYVMTNNHVVAGIEQSGVLEVSFSDGTTAQADFIAGDPSTDIAYIK 194

178 IDATELPVYKIGNPKULKKEGWVAAIAGAPGFDSVTAGIVSAKGRSL-----PNEGY 231
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 IDVSNLPVMSFGDSALGVGQSVMAVGSPLGLSSTVTGIVSAVNRPVRAAGDGESSL 254

232 TPIOTDVAINPQNSGGLFNLKQGVVINSQI-----YSRSGFGMISFAIPIDVAMV 286
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 IDAIOTDAIINVNGSGPLVNDMDGNLIGMNSVIVASISSTSDSAGSIGLGFISPNFAKRV 314

287 AEOIKNTGKVQRGQLGVIIQEVSYGLAQSFGLDKA--SGALLAKILPGSPAERAGLQAGDI 345
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 ADQLISTGVQVTOPIMGVQV-----GTDNSVTGAVIASVQDGGPAAADAGLQPGDI 363

346 VLSLDGEIRSSGDLPMVWGAITPGEVSLGV 377
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 VTKLNDRVIDSPSLIAAARSHDFGETVTLTI 395

RESULT 13
; US-09-712-363-182
; Sequence 182, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182
; LENGTH: 464
```

```

; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-182

Query Match      17.8%; Score 418; DB 9; Length 464;
Best Local Similarity 32.5%; Pred. No. 1.6e-25;
Matches 116; Conservative 60; Mismatches 139; Indels 42; Gaps 10;

Cy 67 AVNNIQAAPAPRTQNGSGNAETDSDPLADSDP--FYEFFKALVPMNPEIQEABDGGUN 124
Db 122 SLVGNRRAPAGSGSGPVAAASAPSTPAAMPPGVSQVAKVPSTVML---ETDGRGS 178
Cy 125 -FGSGPITSKNGYILNTHTVAG---MGS---IKVLNDRREYAKLIGSDVSDVA 174
Db 179 EESGIIISAEGLITNNHVIATAAKPPLGPPKTTTFSDRTAPFTVAGADPISDIA 238
Cy 175 LKIDATELPPVKIKGNPKLKGEMVAALGAPFGFDSNVTAGIVAKGRSLP----- 227
Db 239 VVRVGVSGLPPLISLGGSSDDLVRVQFVLAIISPLGLEGTVTGIVSALNRPVSTGEAGN 238
Cy 228 NESYPPFIQTDAINPGNSGGLFNLKGQVNGINSQIYS-----RSGGFMGISFAIRP 280
Db 239 QNTVDAIQTDPAITPGNSGGLVNMADLVGNNAIATLGADSAQSGSISLGRPIIV 358
Cy 281 DVANNVAQKNTGKVRQGLGVIIQEVSYGLAQSFGLPKAS-GALIAKILPGSPARAG 339
Db 359 DQAKRIADELISGTAKASHASLGVQTN-----DKDTLGAKIVEVAGGAANAG 407
Cy 340 LQAGDIVLSLDGEIRSSGDLFVPMGALTPGKEVSLGCVN-RKGEETITRAKLCNAAE 395
Db 408 VPKGVVTVKVDPRPINSADALVAASVSKAPGATVALTFODPGSGSRVTVQLCKAEQ 464

RESULT 14
US-10-197-634-1
; Sequence 1, Application US/10197634
; Publication No. US20030073629A1
; GENERAL INFORMATION:
; APPLICANT: Alnemi, Emad S.
; TITLE OF INVENTION: OM1 AND DOMAINS THEREOF THAT DISRUPT
; FILE REFERENCE: 480140.479
; CURRENT APPLICATION NUMBER: US/10/197,634
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-197-634-1

Query Match      17.6%; Score 414.5; DB 9; Length 458;
Best Local Similarity 31.8%; Pred. No. 3e-25;
Matches 110; Conservative 65; Mismatches 108; Indels 63; Gaps 10;

Cy 58 AQLVSEBPAAVNIQAAPARTQNGSGNAETDSDPLADSDPFEFFKALVPMNPEIQE 117
Db 151 ADVVEKTAAPVYIIE-----ILDRHP---FLGREVP----- 178
Cy 118 ADDGGLNFGSGPILSNNGYILNTHTVAGMSIKVLNDRREYAKLIGSDVSDVALK 177
Db 179 -----ISNGSGVVAADGLIVNAHVADRNRVRLVLSGDTYEAVTVADVPADIAITLR 233
Cy 178 IDATELPPVKIKGNPKLKGEMVAALGAPFGFDSNVTAGIYS-----AKGRSLPESYT 232
Db 234 IQTKPLPTLPLGRSADVAGQGFVVMGSPFALQNTITSIGIVSSAQRPADILP-QTNV 292
Cy 233 PEIQDVAINPNSGGLFNLKGQVNGINSQIYSRSGFMGISFAIRPIDVANNVAE--QL 230
Db 293 EYIQTDAAIDFNSGGLVNLDEGVGVNMYKT-----AGISFALPSDLRPLRGEK 347
Cy 291 KNTGK----VQRGQLGVIIQEVSYGLAQSFGL-----DKASGALIAKILPGSPARAGL 340
```

```

Db 348 KNSSGISGSGRRYIGVMMTLTSPSILAEQLRPSPEPDVQHGVLHKKVILGSPHRAGL 407
Cy 341 QAGDIVLSLDGEIRSSGDLFVPMGALTPGKEVSLGCVNRKGEETITR 386
Db 408 RPDVILAIIGEQWQNAED--VEAVFTQSGLAVQIRGRETTLT 450

RESULT 15
US-09-796-753-34
; Sequence 34, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 34
; LENGTH: 348
```


TYPE: PRT
ORGANISM: Mouse
US-09-796-753-34

Query Match 17.6%; Score 414; DB 9; Length 348;

Best Local Similarity 33.7%; Pred. No. 2.3e-25;
Matches 112; Conservative 58; Mismatches 98; Indels 64; Gaps 10;

```
QY 58 AQLVQSEGPVAVNIOAPAPRTONGSGNAETDSDPLADSDPF--YEPFKRLVPMNPEIIPQ 115
DB 39 ADVVEKIAPAVHIEL-----FLRHPLFGKNVP-----66
QY 116 EEADGGILNFGSGFIISKNGYILTNTHVA-----GMGSIKVLLNDREYTAKLIGSDV 169
DB 67 -----LSSGSGFIMSEAGLITNAHVSSNSAAPGRQOLKVQLONGDSYEATIKDIDK 119
QY 170 QSDVALLKIDATEELPVVKIGNRKNLKGEMVAIGAPFGFDSVTAGIVSA--KGRSL 226
DB 120 KSDIATIKIHPKKKLPVLLGHSADLRPGEFVVAIGSPFALQNTVTGTIVSTAOREGREL 179
QY 227 P-NESTYPIQTDVAINPGNSGGPLFNLKGQVVGINSQIYRSGGFMGISFAIPID-VAM 284
DB 180 GLRDSMDYIQTDAINYNSGGPLVNLDSGVIGINTLKT-----AGISPAIPSDRITR 234
QY 285 NVAE-QLKNTGKVQRQOLGYLIEVSYGLAQSFGL-----DKASGALIAKILPGSPAER 337
DB 235 FLTEFDQKQIKDWKKRFIGIRMRITITPSLVDELKASNPDPPEVSSGIYVGEVAPNSPSQR 294
QY 338 AGLQAGDIYLSLDGGEIRSSGDLPMWGAITP 369
DB 295 GGIQDGDIIYKVGGRPLVDSSELQEAVALTESP 326
```

Search completed: July 11, 2003, 10:42:52
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:28:37 ; Search time 42 Seconds
(without alignments)
1064.345 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKKYQYFALALCAALLAG.....BRAGLRHGEDELLAVRASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Maximum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2246	95.5	499	2 B81914	probable periplasm
2	985.5	41.9	474	2 P83550	serine proteinase
3	862.5	36.7	514	2 A82581	periplasmic protei
4	735	31.2	523	2 A97479	probable serine pr
5	735	31.2	523	2 A12696	serine proteinase
6	716	30.4	493	2 C97605	probable serine pr
7	716	30.4	514	2 AE2827	serine proteinase
8	693.5	29.5	513	2 AD3418	proteinase DO (EC
9	692.5	29.4	513	2 I40060	serine proteinase
10	683	29.0	497	2 B81728	serine proteinase,
11	683	29.0	524	2 AG3328	probable do serine
12	683	29.0	524	2 AG3328	proteinase do (EC
13	669.5	28.5	488	2 G81528	serine proteinase,
14	668.5	28.4	488	2 H86112	DO serine proteina
15	668.5	28.4	488	2 H86112	DO serine proteina
16	653.5	27.8	457	2 AG0413	proteinase (EC 3.4
17	649	27.6	481	2 AH0410	global stress regu
18	648	27.5	466	1 A64113	heat shock protein
19	639	27.2	474	2 A13349	proteinase DO (EC
20	638	27.1	511	2 P77720	periplasmic serine
21	637	27.1	474	1 I40059	htrA-like protein
22	636.5	27.1	475	2 AC0528	protease DO precu
23	636	27.0	530	2 P87590	heat shock protein
24	633.5	26.9	475	1 S15337	serine proteinase
25	632.5	26.9	456	2 P82307	proteinase DO VC05
26	631	26.8	429	2 AD1894	serine proteinase
27	627.5	26.7	455	2 UC6051	trypsin-like prote
28	623.5	26.5	455	2 C91142	serine endoprotein
29	617.5	26.2	455	2 P85987	serine endoprotein

30	617.5	26.2	455	2 AB0909	serine protease (E
31	615.5	26.2	513	2 B71722	probable periplasm
32	614.5	26.1	459	2 P72359	periplasmic serine
33	610.5	25.9	474	2 S45229	proteinase DO (EC
34	610.5	25.9	474	2 E85500	proteinase DO (EC
35	610.5	25.9	474	2 E90649	proteinase DO (EC
36	606.5	25.8	452	2 S77538	serine proteinase
37	604.5	25.7	481	2 D82826	heat shock protein
38	600.5	25.5	416	2 S75445	proteinase htrA (E
39	594.5	25.3	478	2 G84956	proteinase do prec
40	589	25.0	389	2 B83089	AlgW protein PA444
41	582	24.7	476	2 H71936	proteinase DO - He
42	574.5	24.4	443	1 C64647	serine proteinase
43	571	24.3	453	1 B70426	periplasmic serine
44	564	24.0	468	2 A12811	serine proteinase
45	564	24.0	495	2 B97590	htrA protein homol

ALIGNMENTS

RESULT 1									
B81914									
probable periplasmic serine proteinase (EC 3.4.21.-) NMA0710 [imported] - Neisseria men									
C:Species: Neisseria meningitidis									
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001									
C:Accession: B81914									
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Chumcher, C.; Klee, S.R.; Morel]									
; Holroyd, S.; Dageles, K.; Leather, S.; Moulle, S.; Mungall, K.; Quail, M.A.; Rajadream,									
Nature 404, 502-506, 2000									
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.									
A:Reference number: A81775; PMID:20222556; PMID:10761919									
A:Accession: B81914									
A>Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-499 <PAR>									
A:Cross-references: GB:AL162754; GB:AL157959; NID:97379424; PIDN:CAB83996.1; PID:97379434									
A:Genetic source: serogroup A, strain Z2491									
C:Genetics:									
A:Gene: NMA0710									
C:Superfamily: Helicobacter serine proteinase									
C:Keywords: hydrolase; serine proteinase									
Query Match 95.5%; Score 2246; DB 2; Length 499;									
Best Local Similarity 95.9%; Pred. No. 3,26-138;									
Matches 444; Conservative 6; Mismatches 13; Indels 0; Gaps 0;									
QY	1	VFKKYQYFALALCAALLACERAGSPFGADKKKASFEVERIEHTKDDGSVMLLPDFAQL	60						
DB	1	MFKKYQYFALALCAALLACERAGSPFGADKKKASFEVERIEHTKDDGSVMLLPDFAQL	60						
QY	61	VOSEGPVAVNIIQAAPARTONGSGNAETDSDPLADSPFFYFFFRKVLVPMMEIPEEADD	120						
DB	61	VOSEGPVAVNIIQAAPARTONGSGNAETDSDPLADSPFFYFFFRKVLVPMMEIPEEADD	120						
QY	121	GGNFGSGFIIISKNGVILTNTHVAVAGWSIKVLNDRKREYAKLIGSDVQSDVALLKIDA	180						
DB	121	GGNFGSGFIIISKNGVILTNTHVAVAGWSIKVLNDRKREYAKLIGSDVQSDVALLKIDA	180						
QY	181	TEELPVVKIKNPKNLKGEWVAIIGAPFGFVNSVTAGIVSAKGSLEPNESYTPFIQTDA	240						
DB	181	TEELPVVKIKNPKNLKGEWVAIIGAPFGFVNSVTAGIVSAKGSLEPNESYTPFIQTDA	240						
QY	241	INFGNSGGPLFNLKGVVNGINSQIYSSGGFMGIGSFAIPIDVAVNVAVQLKNTGVVQGG	300						
DB	241	INFGNSGGPLFNLKGVVNGINSQIYSSGGFMGIGSFAIPIDVAVNVAVQLKNTGVVQGG	300						
QY	301	LGVTIOEVSYGLAOSFELDKASGALIKILPGSPAERAGLAQGIIVSLDGEIRSSGDL	360						
DB	301	LGVTIOEVSYGLAOSFELDKASGALIKILPGSPAERAGLAQGIIVSLDGEIRSSGDL	360						
QY	361	PVAVGAIITPGKEVSLGWRKGEIITIKAKLGNAAEHTGASSKTDEAPYTEOQSGTFVES	420						
DB	361	PVAVGAIITPGKEVSLGWRKGEIITIKAKLGNAAEHTGASSKTDEAPYTEOQSGTFVES	420						

Db 361 FVMWGAIIPGKEVSLGVMRKGEIITTKVLLGNAAEHIGASSKTDDEAPYTEQSGSTSVES 420

Qy 421 AGITLQHTDSSGKHLVVVRVSDAAERAGLRGDELLAVRASP 463

Dd 421 AGITLQHTDSSGKHLVVVRVSDAAERAGLRGDELLAQVNP 463

RESULT 2

serine proteinase MucD precursor PA0766 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C/Species: *Pseudomonas aeruginosa*
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 03-Jun-2002
C/Accession: F83550
R/Owner: C.K.; Pham, X.Q.; Ewlin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
Reference number: AB2950; MUID:2043737; PMID:10984043
Accession: F83550

A;Cross-references: GB:AE004511, GB:AE004091, NID:g93946646, PIDN:AAG04155.1, GSPDB:GN001001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: mucD, PA0766
C;Superfamily: Helicobacter serine proteinase

Query Match	41.9%	Score 985.5	DB 2	Length 474
Best Local Similarity	52.5%	Pred. No. 1.8e-56		
Matches 220, Conservative	55	Mismatches 111	Indels 33	Gaps 11

Oy 54 LPDPAQLVSGSPAVVNI-----QAAPAPRTQNGSGNAETSDSLDADSDP-FYEFEFKRLVNP 109
 Db 27 LPDTPPLVEASPAVVNISTRÖKLP---DRAMARGLSIPDLLEGLPMPFRFLERSTQPO 82

110 MPEIP---QEADDDGGLFGSGFIISKGYILTNTHVAVAGMSIKVLLNDKREYAKLIG 166

03 VFKNFKQKQKBAQ ---SUSSGF LLSNDGJ LELNNHVVAUADJELUVKLSKSKSENNKXLLIG L38

Db 139 ADPRSDVALEKTEA-KNLPTKLGDSNKIKVGEWYLATGSPGEEDHSVTAGIVSAKGRSL 197

227 PNESYTPFIQTDVAINPGNSGGPLFNLKGVVGINSOIYSRSGGEMGISFAIPIDVAMNV 286

198 PNESYVPIQIDVAINPGNSGPELNLQGEVGINSQIFTRSGGINGLSFAIPIDVALNV 257

258 ADOLKKAKKISRBGWTGWVTOFNKNDLASEGCI DKPSGAIYNAOIYENGPAKGGIOWENVT 317Z

347 LSLDGEIRSSGDLFVMVGAITPCKEVS LGVWRKGEETIKAKLGNAEHTGASSKTDE - 405

Db 318 LSLNGQINESADLPHLVGNMKGDKINDVIRNGQRKSLMAVGSLPD-----DDEE 370

[illegible]

RESULT 3

periplasmic protease Xp2241 (imported) - *Xylella fastidiosa* (strain 945c)
C.Species: *Xylella fastidiosa*
C.Date: 18-Aug-2000 #sequence_rev1stn 20-Aug-2000 #ext_change 02-Sep-2000

A: Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
Nature 406, 151-157, 2000

A;Note: for a complete list of authors see reference number A59328 below

A:Accession: A82581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <STM>

A: C1088-1; References: GB:AEU0403/1; GB:AEU038493; NID:921013594; ELDN:AAE85040.1; GSPUB:G000047; Experimental source: strain 9.5C
R:Stimson, A.U.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A) Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carret, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.U.S.
submitted to GenBank June 2000

Submitted to *Cell* on October 1, 2004; accepted for publication on November 10, 2004.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address correspondence and reprint requests to Dr. J. A. Rothman at the Department of Cell Biology and Physiology, University of Maryland School of Medicine, 725 North Wolfe Street, Baltimore, MD 21201 (e-mail: jrothman@ummsystem.edu).

© 2005 American Society for Cell Biology 0021-9525/05/251111-11\$15.00/0

DOI: 10.1093/jcellbio/jci325

Keywords: cell cycle; cyclin; cyclin-dependent kinase; mitosis; protein kinase C; transcription factor

Abstract The cyclin-dependent kinase (CDK) 2 is a central component of the cell cycle machinery. CDK2 is regulated by phosphorylation and by association with cyclins. We have identified a novel, ubiquitously expressed, 11-kDa protein, *CDK2AP1*, that interacts with CDK2. *CDK2AP1* is a member of the *CDK2AP* family, which includes *CDK2AP2* and *CDK2AP3*. *CDK2AP1* is a nuclear protein that is enriched in the nucleus and is involved in the regulation of CDK2 activity. *CDK2AP1* is a member of the *CDK2AP* family, which includes *CDK2AP2* and *CDK2AP3*. *CDK2AP1* is a nuclear protein that is enriched in the nucleus and is involved in the regulation of CDK2 activity. *CDK2AP1* is a member of the *CDK2AP* family, which includes *CDK2AP2* and *CDK2AP3*. *CDK2AP1* is a nuclear protein that is enriched in the nucleus and is involved in the regulation of CDK2 activity.

F. G.; Nunes, M. A.; Oliveira, M. C.; de Oliveira, R. C.; Palmieri, D. A.; Rodrigues, V.; Rosa, A. J. de M.; de Rosa Jr., V. E.; de Sa, R. G.; Santelli, R. V.; Sawasaki, A.; Authors: da Silva, A. C. R.; da Silva, F. R.; da Silva, A. M.; Silva Jr., W. A.; da Silveira, M.; Tsuchiko, M. H.; Vallada, H.; Van Sluys, M. A.; Verjovski-Almeida, S.; Vettore, A. L.; Zerbini, A.; Reference number: A59328

A;Gene: XF2241
C;Superfamily: Helicobacter serine proteinase

Query Match	36.7%	Score 862.5;	DB 2;	Length 514;
Best Local Similarity	46.2%;	Pred. No. 1,9e-48;		
Matches 194; Conservative	64;	Mismatches 129;	Indels 33;	Gaps 10;

QY 54 LPDFAQLVQSEGPVVNIQAAPAPRTQNGSGNAETDSDPLADDPFYEFKRL-----V 107

00- 108 DMDBETDPEFADDCG---NEGCGETSKNVCYTNTTNTVACMGSTKVIINDKREVTART. 16A

Db 107 P N Q P R G G Q D - D E G G I A R G M G S G F I I S K D G Y I L T N H H V I T G A S E V Y I K L T D R R E F F K A K I 164

165 IGSDDVQSDVALLKIDATEBEPVKIKGNPKLKPGEWVAIGAPEGFDSVTAGIVSAKGR 224

[illegible]

Db 224 STSDDRYVFICTDVPINQNSGGPLNTRGEVIGINSQIFASAGGYMGISFAIPNLA 283

QY 284 MNVAEQLKNTGKVQRGQLSVIIQEVSYGLAQSGFLDKASGALLIAKILPGSPAERAGLQAG 343

DB 284 INAAEQLKAIKGVKRSMLGVEIGFIDALIKVQIGLGLPUSRGALLVNNIPFHSFAAKAGLEVG 343

Db 344 DVIRSVNGKVISSFSDIPLIGMPGKTATGIRDGKPREIVTTLALN-ODATASED 402

402 KTDEAPYTEQSGTFSVASGITLTQHTDSSGKHLVVRVSDAEPAGLRHDEILAVRA 461

Db 403 EDDYAPSKPETA--INVELDGLQVENTLSAABERLT-----ASSQNAKGGVRIYAVIA 452

A97479

CISpecies: Agrobacterium tumefaciens
CIdate: 30-Sep-2001 #sequence_revision 30-Sep-2001 #ext_change 17-May-2002
CAccession: A37479
RAccession: B. Hinkle, G. Gattung, S. Miller, N. Blanchard, M. Quirello, B. Goldman, R. Goodner.

Science 294, 2323-2328, 2001

A;Reference number: A97359; PMID:11743194
A;Accession: A97479

A: Molecule type: DNA
A: Position: 1-523, 471P

A/CIOBS-REFERENCES: GB:AEUV/003/ FIDN:MM00/00.1/ FID:JGJ0155300/ USFUD:GNOV0103

C:Genetics:
 A:Gene: AGR_C_1792
 A:Map position: circular chromosome
 C:Superfamily: Helicobacter serine proteinase

Query Match 31.2%; Score 735; DB 2; Length 523;
 Best Local Similarity 37.0%; Pred. No. 3.7e-40;

Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;

QY 36 SFVERIETKDDGVSMLLPDPAQLVQSEGPVAVNIQAAPARTONGSNAETDS-----90

DB 44 SFAAPVETAPQ-----VPSFANVVDAVSPAVSV-----RVQSNVQPADSDSNFSF 91

QY 91 -----DPLADSDPFYEFKRLVPMNPBIPQSEADGGLN-----FGSGFI 130

DB 92 NFGGRGLDQLPDHPLKRFKFEFGQ-----NQDRSDGPRNRHDKGPLRPVAGSGGF 146

QY 131 ISKNGYILTNTHVAVAGMSIKVLNDRREYTKLIGSDVQSDVALLKIDATEELPVYKIG 190

DB 147 ISEGGYVVTNNHVVDDGSAATVVMNDGTELEKALVGRDPRITLALKVDVNRKFTYVYKFA 206

QY 191 NPNKLRGGEVVAALGAPFGFNSVTAGIVSAKGRSLPNESYTPFIQTDAINPNSGGPL 250

DB 207 DDTKIRVGDVWVAAGNPGFGLGTVTSGLISARGRDISGPRVDYQLDAAVNRKNSGGPA 266

QY 251 FNLKGQVVGINSQIYSSSGFMGISFAIPIDVAMNVAEQLNKTKVQRGOLGVIIQEVSY 310

DB 267 FNLNGEVVGINTALFSPSGGNVGIAPFIPSSVAKVIALDLQKDGKVERGWLGVQIOPYSK 326

QY 311 GLAQSGFLDKASGALITAKILPGSPAERAGLQAGDIVLSLDGEIRSSGDLPMVGAITPG 370

DB 327 DIAESLGLAEAKGALVVSPOSGSPDKAGIKQGDITITVANGDPVARDLSRRIGMAPN 386

QY 371 KEVSLGVWRKGEETITKAKLGN--AAEHTGASSKTDAPYTEQSGTFVSASGITLQTH 428

DB 387 SKYEISLWRGKSGSVTVTLGDLTSDASAKATPSONDKGSGSSSEKVLSTGLTVSPS 446

QY 429 TDSGSKHLVVVR---SDAERAGLRHGDILLAV 459

DB 447 DDGNG--LAITVDVDPDSDAAR-GLTKGEKITSV 477

RESULT 5
 A12696
 serine proteinase DO-like proteinase dop [imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 17-May-2002
 C:Accession: A12696
 R:Wood, D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 P., G.; Giller, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; Mclellan,
 ster, B.W.
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: A12696
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-523 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAI4191.1; PID:g17739363; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: dop
 A:Map position: circular chromosome
 C:Superfamily: Helicobacter serine proteinase

Query Match 31.2%; Score 735; DB 2; Length 523;
 Best Local Similarity 37.0%; Pred. No. 3.7e-40;
 Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;

QY 36 SFVERIETKDDGVSMLLPDPAQLVQSEGPVAVNIQAAPARTONGSNAETDS-----90

DB 44 SFAAPVETAPQ-----VPSFANVVDAVSPAVSV-----RVQSNVQPADSDSNFSF 91

QY 91 -----DPLADSDPFYEFKRLVPMNPBIPQSEADGGLN-----FGSGFI 130

DB 92 NFGGRGLDQLPDHPLKRFKFEFGQ-----NQDRSDGPRNRHDKGPLRPVAGSGGF 146

QY 131 ISKNGYILTNTHVAVAGMSIKVLNDRREYTKLIGSDVQSDVALLKIDATEELPVYKIG 190

DB 147 ISEGGYVVTNNHVVDDGSAATVVMNDGTELEKALVGRDPRITLALKVDVNRKFTYVYKFA 206

QY 191 NPNKLRGGEVVAALGAPFGFNSVTAGIVSAKGRSLPNESYTPFIQTDAINPNSGGPL 250

DB 207 DDTKIRVGDVWVAAGNPGFGLGTVTSGLISARGRDISGPRVDYQLDAAVNRKNSGGPA 266

QY 251 FNLKGQVVGINSQIYSSSGFMGISFAIPIDVAMNVAEQLNKTKVQRGOLGVIIQEVSY 310

DB 267 FNLNGEVVGINTALFSPSGGNVGIAPFIPSSVAKVIALDLQKDGKVERGWLGVQIOPYSK 326

QY 311 GLAQSGFLDKASGALITAKILPGSPAERAGLQAGDIVLSLDGEIRSSGDLPMVGAITPG 370

DB 327 DIAESLGLAEAKGALVVSPOSGSPDKAGIKQGDITITVANGDPVARDLSRRIGMAPN 386

QY 371 KEVSLGVWRKGEETITKAKLGN--AAEHTGASSKTDAPYTEQSGTFVSASGITLQTH 428

DB 387 SKYEISLWRGKSGSVTVTLGDLTSDASAKATPSONDKGSGSSSEKVLSTGLTVSPS 446

QY 429 TDSGSKHLVVVR---SDAERAGLRHGDILLAV 459

DB 447 DDGNG--LAITVDVDPDSDAAR-GLTKGEKITSV 477

RESULT 6
 C97605
 probable serine proteinase DO-like precursor [imported] - Agrobacterium tumefaciens (stre
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: C97605
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
 A:Reference number: A97359; PMID:11743194
 A:Accession: C97605
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-493 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK87796.1; PID:g15157169; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_3700
 A:Map position: circular chromosome

Query Match 30.4%; Score 716; DB 2; Length 493;
 Best Local Similarity 36.4%; Pred. No. 5.8e-39;
 Matches 170; Conservative 75; Mismatches 146; Indels 76; Gaps 10;

QY 48 GSVSMILPDPAQLVQSEGP-----AVVNIQAAPARTONGSNAETDSPLAD 95

DB 3 GSLS---APVABARQSHGPEVSADLAERPLDVAVINISTQNKTE-CKEPV---PKLRPE 55

QY 96 SDPFYEFKRLVPMNPBIPQSEADGG---LNFSGFIISKNGYILTNTHVAVAGMSIKY 152

DB 56 GSPFOEFER-----DYFDSQKPEGGEKVNLSGSGFVIDPAGYVVTNNHVIIGADAIEV 108

QY 153 LINDREYTKLIGSDVQSDVALLKIDATEELPVYKIKPKNKLPGEMVAALGAPGFND 212

DB 109 IFPNQSKIKATLVGDTITDLSVLKVEKPTPLKAYKFGDSRNRIGDWMAVAGNPGGLG 168

QY 213 SVTAVIVSAKGRSLPNESYTPFIQTDAINPNSGGPLFNLKGQVVGINSQIYSSSGFM 272

DB 169 SLTVGVISARGNINAGPYDNIQTDAIINKNSGGPLFNMKGVEVIGINTAIISPGGSI 228

QY 273 GISFAIPIDVAMNVAEQLNKTKVQRGOLGVIIQEVSYGLAQSGFLDKASGALITAKILPG 332

Db 229 GIGFAVPTLQNIIVQQLIEFGETRGMLGVROPVTDVVAASIGMSAKALISGVAKG 288

Qy 333 SPAERAGLQAGDIVLSLDGGEIRSGDLPMVVGAIITPGEKVSIGWRKGEITIKAKLGN 392

Db 289 GFVENGPIQAGDVVLFKDGKINEMRDLRIIVASPVGKEVDVVYVKGKEITIKVLLGQ 348

Qy 393 AAEHTGASSKTDEAPYTEQSG-----TFVSESAGITLQTHDSSGK 434

Db 349 LQDTTDEKASTDD-PCGEGDDGSMVAAPDDKGGDDAQDQTPPEVREAPOTV-----LGM 401

Qy 435 HLIVV-----VRVSDAERAGLRHGDDELAV 459

Db 402 NLVVLNSNELRTKGIASVEGVLVASVDPSPAEQKMKAGDIVVEV 448

RESULT 7
AE2827
serine proteinase DO-like precursor htra (imported) - Agrobacterium tumefaciens (strain
Accession: AE2827
Species: Agrobacterium tumefaciens
Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
Accession: AE2827
R:Wood, D.W.; Saebul, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I.
erage, G.; Gille, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell
: Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
A:er, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577, PMID:11743193
A:Accession: AE2827
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43035.1; PID:G17740500; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: htra
A:Map position: circular chromosome

Query Match 30.4%; Score 716; DB 2; Length 514;
Best Local Similarity 36.4%; Pred. No. 6,1e-39;
Matches 170; Conservative 75; Mismatches 146; Indels 76; Gaps 10;

Qy 48 GGSVMLPDPFAQLVQSEGP-----AVVNIQAPAPRTONGSGNAETDSDPLAD 95

Db 24 GSLT---APVAAQAQHGPEVSADLAEPFLDAAVNNISTSQNVKTE-CKGPV---PRLTPE 76

Qy 96 SPPFEFFFKRLVPMMEIIOEADDCG---LNFSGGFIISKNGYILNTHTVVAAGMSIKY 152

Db 77 GSPFOEFFK-----DYFDSQKPEEGEKVNSLGSQFVIDPAGVYVNNHVIIEGADALEV 129

Qy 153 LINDKREYFAKLIGSVOSDVALIKTIDATELFPVVKGNPKLKGEMVVAIGAEPFDN 212

Db 130 IFPNSSKLKATLVGTITKTDLSTLVKBPKPLKAVKFGDSRMRIGMWVAAGNPFLGG 189

Qy 213 SVTAGIVSAGKSLPNESTYTPITQDVAINPNSGGFLFNLKQGVVNGINSQIVSRSGFM 272

Db 190 SILTVGISARGRINAGPYDNFIQTDALINKNSGGFLFNMKGCEVICINNAIISPSSGI 249

Qy 273 GISFAPIIVANVVAQLKTKGVQNGQLVITIQEVSYGLAQSGFLDKASGALIAKLPG 332

Db 250 GIGFAVPTLQNIIVQQLIEFGETRGMLGVROPVTDVVAASIGMSAKALISGVAKG 309

Qy 333 SPAERAGLQAGDIVLSLDGGEIRSGDLPMVVGAIITPGEKVSIGWRKGEITIKAKLGN 392

Db 310 GEVENGPIDAGDVVLFKDGKINEMRDLRIIVASPVGKEVDVVYVKGKEITIKVLLGQ 369

Qy 393 AAEHTGASSKTDEAPYTEQSG-----TFVSESAGITLQTHDSSGK 434

Db 370 LQDTTDEKASTDD-PCGEGDDGSMVAAPDDKGGDDAQDQTPPEVREAPOTV-----LGM 422

Qy 435 HLIVV-----VRVSDAERAGLRHGDDELAV 459

Db 423 NLVVLNSNELRTKGIASVEGVLVASVDPSPAEQKMKAGDIVVEV 469

RESULT 8
AD3418
proteinnase DO (EC 3.4.21.-) (imported) - Brucella melitensis (strain 16M)
Accession: AD3418
Species: Brucella melitensis
Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
Accession: AD3418
R:DeIvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.
: Mazur, M.; Goldeman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leteac
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
Accession: AD3252; PMID:11756688
A:Accession: AD3418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-513 <KUR>
A:Cross-references: GB:AE00917; PIDN:AAL52511.1; PID:G17983322; GSPDB:GN00190
A:Experimental source: strain 16M
A:Genetics:
A:Gene: BME11330
A:Map position: I
A:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 29.5%; Score 693.5; DB 2; Length 513;
Best Local Similarity 37.4%; Pred. No. 1,8e-37;
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

Qy 9 AALCALLAGCEKASFFGADKKEASFVERIEHTKDGVSMLLPD---FAQLVQSEG 65

Db 14 AAVALSAL-----AGAFVVTGFLGALNEMRAE-----AVHTVPPQAGADPDLVEKRV 61

Qy 66 PAVVNIQAPAPRTONGSGN---AETDSDPLADSPFEFF-----KRLV 107

Db 62 PAVSVRVKKDVQETSNRGPFGPPGPDQLPDGHLRFFRDFGEMPRGDSRSDNRK 121

Qy 108 PMMEIIOEADDCGLNFGSGFIISKNGYILNTHTVVAAGMSIKYLLINDKREYFAKLIGS 167

Db 122 ANKRPPGHERP-----VAQSGFVISEDDVYVNNHVSDDGAYTVVLDGDTLDAKLTA 177

Qy 168 DVQSDVALIKTIDATE-ELPVYKGNPKLKGEMVVAIGAEPFNSVTAGIVSAGKSL 226

Db 178 DPTDVLAKINAKRKRVVYAFGDNDKRVGDWVAAGNPFLGGLGTVTSIGVSARGDI 237

Qy 227 PNESTYTPITQDVAINPNSGGFLFNLKQGVVNGINSQIVSRSGFMISFAPIIVANV 286

Db 238 GAGPYDDPTQIDAAVNKNSSGSPAPFDSGEVIGINTALFSPGSGSVGAIFALPSSTAKOV 297

Qy 287 AEQLKNTGVQNGQLVITIQEVSYGLAQSGFLDKASGALIAKLIPSPAERAGLQAGDIV 346

Db 298 VDQIKKGSVGRGVIQVQIOTVDIAASLGAELKAVASPODDGPAKAGIRAGDVI 357

Qy 347 LSLDGEIRSSGDLPMVVGAIITPGEKVSIGWRKGEITIKAKLGN 403

Db 358 TAVVGETVQDPRDLARKVANIAPGEKALTVWRKKKAEINVTIAMPNDKSGSGQSD 417

Qy 404 DEAPYTEQSGTFVSAGITLQTHDSSGHLIVVRY---SDAERAGLRHGDDELAV 459

Db 418 NDGGQGE-----TLDSGLTVVPSD---GKGVVTVDPDPSDAADR-GIRSGDIVSV 467

RESULT 9
I40060
serine proteinase (EC 3.4.21.-) htra, temperature-inducible [validated] - Brucella abortu
C:Species: Brucella abortus
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 26-May-2000
C:Accession: I40060
R:Tatum, F.M.; Cheville, N.F.; Morfitt, D.
Microb. Pathog. 17, 23-36, 1994
A:Title: Cloning, characterization and construction of htra and htra-like mutants of Bruc
A:Reference number: I40059; MUID:95165990; PMID:7861951

A:Accession: 140060
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-513 <RSS>
A:Cross-references: EMBL:U07352; NID:g497156; PIDN:AAA70164.1; PID:g497157
C:Genetics:
A:Gene: htra
C:Function:
A:Description: EC 3.4.21.-; serine proteinase [validated, MUID:95165990]; permits growth
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 29.4%; Score 692.5; DB 2; Length 513;
Best Local Similarity 37.4%; Pred. No. 26-37; Indels 53; Gaps 13;
Matches 179; Conservative 82; Mismatches 165;

9 ALAALCAALAGCKAGSPFGADKKEASFVEIEHTKDDGSYMLP---DFAQLVQSEG 65
14 AAVALSAL-----AGAFVYTGPGALNEAPAE-----AAHVTPPQAGFDLVKVR 61
66 PAVVNIQAAPAPRTQNGSGN---AETSDPLADSDPEYEF-----KRLV 107
62 PAVVSVRVKDVQGETSNRGPFGPFGPDQLPDGHLKFRFDFGMEPRGDSHSDNRGK 121
108 PNMPELPOEADDCGLAFSGFIIISKNGIYLNTVHVAGKSLKVLINDKREYTKLIGS 167
122 ANMPRPCHERP---VAQSGSFVISEDGVYVNNHVSNGDAVTVLIDGTETLDAKLIGA 177
168 DVQSDVALTKIDATE-ELPVRKIGNPKNLKPEWVAATGAPGFDNSVTAIGVSANGRL 226
178 DPTIDLAVLINAAPKRRKRVYVAFGDDNKRVRQVDWVAANPGLGTVTSGIVSARGRDI 237
227 PNESTPEFIQTDVAINPNSGSGPLFNLKQVVGINSQIYSRSGFWMGISPAIPDIVAMN 286
238 GAQPYDQFIQIDAIVNKGNSGPAFDLGEVIGIMTAITSPPSGSGVIGAFITSSRAKQV 297
287 AEOLKNTGKVGQGLGVITQEVSYGLAOSFGIDKASGALIAKLPGSPAPRAGLQAGDIV 346
298 VPDILIKGSVERGIVGVIQPVTKDIAASIGLAEKGAIVASPODDGPAKAKIGKAGDIV 357
347 LSLDGGIRSSGDLPWVGAITTPGKEVSLGVWR--KGEIITIK-ALGNAHERTGASKT 403
358 TANGETVQDPRIARAKVANIAPGSKAALTWRKKNKAEEINVTIAMPNDRKSGSGSND 417
404 DEAPYTEQSGSTFVESAGITLTQHTDSSGKLIVVVR--SDAARAGLRHGDEITAV 459
418 NDGGQGE-----TLDSYGLTVVPSED--GKGVVTVDDPDSDAADR-GIRSGDIVSV 467

LT 10
28

serine proteinase, Htra/DegQ/Degs family TC0210 [imported] - Chlamydia muridarum (stratified)
C:Species: Chlamydia muridarum, Chlamydia trachomatis Moxn
C:Date: 31-Mar-2000 #sequence revision 31-Mar-2000 #text change 21-Jul-2000
R:Accession: B81728
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Moxn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: B81728
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-497 <RT>
A:Cross-references: GB:AB002286; GB:AB002160; NID:g7190247; PIDN:AAF39082.1; PID:g719022
A:Experimental source: strain N1gg (Moxn)
C:Genetics:
A:Gene: TC0210
C:Superfamily: Helicobacter serine proteinase

Query Match 29.0%; Score 683; DB 2; Length 497;
Best Local Similarity 37.4%; Pred. No. 8-16-37;
Matches 173; Conservative 77; Mismatches 154; Indels 58; Gaps 11;

[illegible]

[illegible]

C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence
C:Accession: AG3328
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Muijer, C.; Los, T.; Ivanova,
I.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesche,
proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the fecundative intracellular pathogen *Brucella melitensis*
A:Accession number: AD5252; PMID:11756688
A:Accession: AG3328
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-524 <KUR>
A:Cross-references: GB:AE008917; PIDN:AL51794.1; PID:q17982537; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME10613
A:Map position: I
A:Keywords: hydrolase; serine proteinase

```

Query Match Similarity   29.0% Score 683; DB 2; Length 524;
Best Local Similarity    40.5%; Pred. No. 8.7e-37;
Matches 161; Conservative 67; Mismatches 152; Indels 18; Gaps 7;

QY      GSVSMLLP-----DFAQLVQSEGPVVNNICAAPRTONGSGNAETSDPLADSDP 98
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      GTVSMCTPALAAOGPASVADLAEGLDAVVNISTSQTVK--DDGEGDGCVPMQPVEGSP 84

QY      FEEFFRLVPNMREIQEEAADGGDLNFSGSFTI--SKNYVILLNTTHYVAAGMSIKVLINLK 157
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      FOEFFDFDF--NDKXGAGGDDSRKYQSLSGSFTIDAEKGIVTNNHVIADADELNVFNNDG 143

QY      REYTAKLIGSDVOVDVALLIKIDATE-ELPVVKIGNPNKLKPGEWYAALGAPEFDSVTYA 216
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      SKLKVELVGKDTITDALILKVDPESKHKLKAHPGISEKARIQDWVLAIONPFGLGCTVTA 203

QY      GTVSAGKGRSLPNESYTPFIOTDVAINPONSNGGPLFNLKGQVGVINSQIYSRSGGFNGISF 276
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      GIISARKRDINSRPYDDPIOTDAAINRGNSCGPLFDMDGKVGIVNTAIIISPFGSGSIGCP 263

QY      AIPIDVAMVMVAEQLKNTKGYORGOGLVIIQEVSGLAOSFGLDKASGALIAKILPPSPA 336
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      AIIPAEMVAAGVIIDDKERGEVRKRGLGVRLCPVTEBIEAIOSLSLKETKGAIIAGLIENSVD 323

QY      RAGLGAGDIVLSLDGGEIRSSGDLPVMVGAITPGKEVSLGWNRKKEBEITIKAKLGAAHEH 396
       ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      NKAIEAGDVIVIRPDGKVPDPTARLDLRIVAERPVGKEVEIIVIRQGAEKTLKYLGLCLVE-- 382

QY      TGASSKTDEAPYTEOOSGTSFSVESAGITLLQHHTDSSK 434
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      ---DDKSTEPAVEDQVPAPDDGEQPGARQET--PKDSDK 416

```

RESULT 13

G81528
serine proteinase, HtrA/DegQ/DegS family CP0877 [imported] - Chlamydothrix pneumoniae (strain H400)
C:Species: Chlamydothrix pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 21-Jul-2000
C:Accession: G81528
R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J.D.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, M.C.; Nierman, W.C.; White, O.; Fraser-Lim, A.; Sutton, G.; Peterson, J.; Smith, T.F.; Venter, A.; Adams, L.M.; Holt, C.; Lapid, A.; Meyer, J.; Nelson, K.; Nelson, W.; Paulsen, O.; Richardson, J.; Salzberg, S.L.; Shumway-Cook, A.; Smith, T.F.; White, O.; Young, L.; Davis, N.; Fouts, D.E.; Gill, S.R.; Hickey, J.D.; Hickey, P.W.; Jaffe, J.; Karp, P.; Karp, M.; Karp, N.; Karp, O.; Karp, P.; Karp, S.; Karp, T.; Karp, U.; Karp, V.; Karp, W.; Karp, X.; Karp, Y.; Karp, Z.; Karp, AA; Karp, AB; Karp, AC; Karp, AD; Karp, AE; Karp, AF; Karp, AG; Karp, AH; Karp, AI; Karp, AJ; Karp, AK; Karp, AL; Karp, AM; Karp, AN; Karp, AO; Karp, AP; Karp, AQ; Karp, AR; Karp, AS; Karp, AT; Karp, AU; Karp, AV; Karp, AW; Karp, AX; Karp, AY; Karp, AZ; Karp, BA; Karp, BB; Karp, BC; Karp, BD; Karp, BE; Karp, BF; Karp, BG; Karp, BH; Karp, BI; Karp, BJ; Karp, BK; Karp, BL; Karp, BM; Karp, BN; Karp, BO; Karp, BP; Karp, BQ; Karp, BR; Karp, BS; Karp, BT; Karp, BU; Karp, BV; Karp, BW; Karp, BX; Karp, BY; Karp, BZ; Karp, CA; Karp, CB; Karp, CC; Karp, CD; Karp, CE; Karp, CF; Karp, CG; Karp, CH; Karp, CI; Karp, CJ; Karp, CK; Karp, CL; Karp, CM; Karp, CN; Karp, CO; Karp, CP; Karp, CQ; Karp, CR; Karp, CS; Karp, CT; Karp, CU; Karp, CV; Karp, CW; Karp, CX; Karp, CY; Karp, CZ; Karp, DA; Karp, DB; Karp, DC; Karp, DD; Karp, DE; Karp, DF; Karp, DG; Karp, DH; Karp, DI; Karp, DJ; Karp, DK; Karp, DL; Karp, DM; Karp, DN; Karp, DO; Karp, DP; Karp, DQ; Karp, DR; Karp, DS; Karp, DT; Karp, DU; Karp, DV; Karp, DW; Karp, DX; Karp, DY; Karp, DZ; Karp, EA; Karp, EB; Karp, EC; Karp, ED; Karp, EE; Karp, EF; Karp, EG; Karp, EH; Karp, EI; Karp, EJ; Karp, EK; Karp, EL; Karp, EM; Karp, EN; Karp, EO; Karp, EP; Karp, EQ; Karp, ER; Karp, ES; Karp, ET; Karp, EU; Karp, EV; Karp, EW; Karp, EX; Karp, EY; Karp, EZ; Karp, FA; Karp, FB; Karp, FC; Karp, FD; Karp, FE; Karp, FF; Karp, FG; Karp, FH; Karp, FI; Karp, FJ; Karp, FK; Karp, FL; Karp, FM; Karp, FN; Karp, FO; Karp, FP; Karp, FQ; Karp, FR; Karp, FS; Karp, FT; Karp, FU; Karp, FV; Karp, FW; Karp, FX; Karp, FY; Karp, FZ; Karp, GA; Karp, GB; Karp, GC; Karp, GD; Karp, GE; Karp, GF; Karp, GG; Karp, GH; Karp, GI; Karp, GJ; Karp, GK; Karp, GL; Karp, GM; Karp, GN; Karp, GO; Karp, GP; Karp, GQ; Karp, GR; Karp, GS; Karp, GT; Karp, GU; Karp, GV; Karp, GW; Karp, GX; Karp, GY; Karp, GZ; Karp, HA; Karp, HB; Karp, HC; Karp, HD; Karp, HE; Karp, HF; Karp, HG; Karp, HH; Karp, HI; Karp, HJ; Karp, HK; Karp, HL; Karp, HM; Karp, HN; Karp, HO; Karp, HP; Karp, HQ; Karp, HR; Karp, HS; Karp, HT; Karp, HU; Karp, HV; Karp, HW; Karp, HX; Karp, HY; Karp, HZ; Karp, IA; Karp, IB; Karp, IC; Karp, ID; Karp, IE; Karp, IF; Karp, IG; Karp, IH; Karp, II; Karp, IJ; Karp, IK; Karp, IL; Karp, IM; Karp, IN; Karp, IO; Karp, IP; Karp, IQ; Karp, IR; Karp, IS; Karp, IT; Karp, IU; Karp, IV; Karp, IW; Karp, IX; Karp, IY; Karp, IZ; Karp, JA; Karp, JB; Karp, JC; Karp, JD; Karp, JE; Karp, JF; Karp, JG; Karp, JH; Karp, JI; Karp, JJ; Karp, JK; Karp, JL; Karp, JM; Karp, JN; Karp, JO; Karp, JP; Karp, JQ; Karp, JR; Karp, JS; Karp, JT; Karp, JU; Karp, JV; Karp, JW; Karp, JX; Karp, JY; Karp, JZ; Karp, KA; Karp, KB; Karp, KC; Karp, KD; Karp, KE; Karp, KF; Karp, KG; Karp, KH; Karp, KI; Karp, KJ; Karp, KK; Karp, KL; Karp, KM; Karp, KN; Karp, KO; Karp, KP; Karp, KQ; Karp, KR; Karp, KS; Karp, KT; Karp, KU; Karp, KV; Karp, KW; Karp, KX; Karp, KY; Karp, KZ; Karp, LA; Karp, LB; Karp, LC; Karp, LD; Karp, LE; Karp, LF; Karp, LG; Karp, LH; Karp, LI; Karp, LJ; Karp, LK; Karp, LL; Karp, LM; Karp, LN; Karp, LO; Karp, LP; Karp, LQ; Karp, LR; Karp, LS; Karp, LT; Karp, LU; Karp, LV; Karp, LW; Karp, LX; Karp, LY; Karp, LZ; Karp, MA; Karp, MB; Karp, MC; Karp, MD; Karp, ME; Karp, MF; Karp, MG; Karp, MH; Karp, MI; Karp, MJ; Karp, MK; Karp, ML; Karp, MM; Karp, MN; Karp, MO; Karp, MP; Karp, MQ; Karp, MR; Karp, MS; Karp, MT; Karp, MU; Karp, MV; Karp, MW; Karp, MX; Karp, MY; Karp, MZ; Karp, NA; Karp, NB; Karp, NC; Karp, ND; Karp, NE; Karp, NF; Karp, NG; Karp, NH; Karp, NI; Karp, NJ; Karp, NK; Karp, NL; Karp, NM; Karp, NN; Karp, NO; Karp, NP; Karp, NQ; Karp, NR; Karp, NS; Karp, NT; Karp, NU; Karp, NV; Karp, NW; Karp, NX; Karp, NY; Karp, NZ; Karp, OA; Karp, OB; Karp, OC; Karp, OD; Karp, OE; Karp, OF; Karp, OG; Karp, OH; Karp, OI; Karp, OJ; Karp, OK; Karp, OL; Karp, OM; Karp, ON; Karp, OO; Karp, OP; Karp, OQ; Karp, OR; Karp, OS; Karp, OT; Karp, OU; Karp, OV; Karp, OW; Karp, OX; Karp, OY; Karp, OZ; Karp, PA; Karp, PB; Karp, PC; Karp, PD; Karp, PE; Karp, PF; Karp, PG; Karp, PH; Karp, PI; Karp, PJ; Karp, PK; Karp, PL; Karp, PM; Karp, PN; Karp, PO; Karp, PP; Karp, PQ; Karp, PR; Karp, PS; Karp, PT; Karp, PU; Karp, PV; Karp, PW; Karp, PX; Karp, PY; Karp, PZ; Karp, QA; Karp, QB; Karp, QC; Karp, QD; Karp, QE; Karp, QF; Karp, QG; Karp, QH; Karp, QI; Karp, QJ; Karp, QK; Karp, QL; Karp, QM; Karp, QN; Karp, QO; Karp, QP; Karp, QQ; Karp, QR; Karp, QS; Karp, QT; Karp, QU; Karp, QV; Karp, QW; Karp, QX; Karp, QY; Karp, QZ; Karp, RA; Karp, RB; Karp, RC; Karp, RD; Karp, RE; Karp, RF; Karp, RG; Karp, RH; Karp, RI; Karp, RJ; Karp, RK; Karp, RL; Karp, RM; Karp, RN; Karp, RO; Karp, RP; Karp, RQ; Karp, RR; Karp, RS; Karp, RT; Karp, RU; Karp, RV; Karp, RW; Karp, RX; Karp, RY; Karp, RZ; Karp, SA; Karp, SB; Karp, SC; Karp, SD; Karp, SE; Karp, SF; Karp, SG; Karp, SH; Karp, SI; Karp, SJ; Karp, SK; Karp, SL; Karp, SM; Karp, SN; Karp, SO; Karp, SP; Karp, SQ; Karp, SR; Karp, SS; Karp, ST; Karp, SU; Karp, SV; Karp, SW; Karp, SX; Karp, SY; Karp, SZ; Karp, TA; Karp, TB; Karp, TC; Karp, TD; Karp, TE; Karp, TF; Karp, TG; Karp, TH; Karp, TI; Karp, TJ; Karp, TK; Karp, TL; Karp, TM; Karp, TN; Karp, TO; Karp, TP; Karp, TQ; Karp, TR; Karp, TS; Karp, TT; Karp, TU; Karp, TV; Karp, TW; Karp, TX; Karp, TY; Karp, TZ; Karp, UA; Karp, UB; Karp, UC; Karp, UD; Karp, UE; Karp, UF; Karp, UG; Karp, UH; Karp, UI; Karp, UJ; Karp, UK; Karp, UL; Karp, UM; Karp, UN; Karp, UO; Karp, UP; Karp, UQ; Karp, UR; Karp, US; Karp, UT; Karp, UU; Karp, UV; Karp, UW; Karp, UX; Karp, UY; Karp, UZ; Karp, VA; Karp, VB; Karp, VC; Karp, VD; Karp, VE; Karp, VF; Karp, VG; Karp, VH; Karp, VI; Karp, VJ; Karp, VK; Karp, VL; Karp, VM; Karp, VN; Karp, VO; Karp, VP; Karp, VQ; Karp, VR; Karp, VS; Karp, VT; Karp, VU; Karp, VV; Karp, VW; Karp, VX; Karp, VY; Karp, VZ; Karp, WA; Karp, WB; Karp, WC; Karp, WD; Karp, WE; Karp, WF; Karp, WG; Karp, WH; Karp, WI; Karp, WJ; Karp, WK; Karp, WL; Karp, WM; Karp, WN; Karp, WO; Karp, WP; Karp, WQ; Karp, WR; Karp, WS; Karp, WT; Karp, WU; Karp, WV; Karp, WW; Karp, WX; Karp, WY; Karp, WZ; Karp, XA; Karp, XB; Karp, XC; Karp, XD; Karp, XE; Karp, XF; Karp, XG; Karp, XH; Karp, XI; Karp, XJ; Karp, XK; Karp, XL; Karp, XM; Karp, XN; Karp, XO; Karp, XP; Karp, XQ; Karp, XR; Karp, XS; Karp, XT; Karp, XU; Karp, XV; Karp, XW; Karp, XX; Karp, XY; Karp, XZ; Karp, YA; Karp, YB; Karp, YC; Karp, YD; Karp, YE; Karp, YF; Karp, YG; Karp, YH; Karp, YI; Karp, YJ; Karp, YK; Karp, YL; Karp, YM; Karp, YN; Karp, YO; Karp, YP; Karp, YQ; Karp, YR; Karp, YS; Karp, YT; Karp, YU; Karp, YV; Karp, YW; Karp, YX; Karp, YY; Karp, YZ; Karp, ZA; Karp, ZB; Karp, ZC; Karp, ZD; Karp, ZE; Karp, ZF; Karp, ZG; Karp, ZH; Karp, ZI; Karp, ZJ; Karp, ZK; Karp, ZL; Karp, ZM; Karp, ZN; Karp, ZO; Karp, ZP; Karp, ZQ; Karp, ZR; Karp, ZS; Karp, ZT; Karp, ZU; Karp, ZV; Karp, ZW; Karp, ZX; Karp, ZY; Karp, ZZ; Karp, AA; Karp, AB; Karp, AC; Karp, AD; Karp, AE; Karp, AF; Karp, AG; Karp, AH; Karp, AI; Karp, AJ; Karp, AK; Karp, AL; Karp, AM; Karp, AN; Karp, AO; Karp, AP; Karp, AQ; Karp, AR; Karp, AS; Karp, AT; Karp, AU; Karp, AV; Karp, AW; Karp, AX; Karp, AY; Karp, AZ; Karp, BA; Karp, BB; Karp, BC; Karp, BD; Karp, BE; Karp, BF; Karp, BG; Karp, BH; Karp, BI; Karp, BJ; Karp, BK; Karp,

C:Superfamily: Helicobacter serine proteinase

Query Match	28.5%	Score 669.5	DB 2	Length 488
-------------	-------	-------------	------	------------

Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

QY 10 LAALCAALLAGCEKAGSFFGADKKEASFEVERIEHTKWDGVSMLLPDFAQLVQSEGPVV 69

Db 10 LAVLVGSXLLALPLSGQAVG--KKE-----RVSELPQDVLKEISGGFSKVATKATPAVV 63

QY 70 NIOAPAPR--TONGSNAETDSDPLADSDPY-EFFKRL--VPNMPEIPQ-BEADDGGL 123

Db 64 YIESFPKSQAVTHPSPGRRGPYENPF--DYFNDEFNRFGLPSQREKPKQSKAVR--- 117

124 NFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSDVQSDVALLKIDATEE 183

Db 118 --GTGFLVSPDGYIVTNNHVEDTGKIHVTLHDGOKYPATVIGLDPKTDLAVIKI-KSON 174

184 L.PVVKI^{GNPKNI}.KPGEWVA^{ATGA}PEFGF^{DNSVT}AGIVSAKGRS-^I.PNE^{SYT}PETIOTDVAIN 242

175 I PYI:SEGNSDHI:KVGDWATA:IGNPEGI:OATVTVGYI:SAKGRNOI:HTIANDEFEI:OTDAATN 234

343 PONECCBI ENT KCVUHCINEATVSBCCCEMCTCEATBDIVAMUUAFOI KUTCVI/ABCIOIC 303

[illegible][illegible]

1. The first part of the document is a list of references. The references are listed in a standard format, with the author's name, the title of the work, and the publisher. The references are as follows:

.....

[illegible][illegible][illegible]

DU 399 IKVQNPFIPIANLWGIAPFINGIPI ISVERPSVHASSGIAPGULIEMV 470

RESULT 14

H86612
DO serine proteinase [imported] - Chlamydothrix pneumoniae (strain J138)

C/Species: Chlamydomonadales, Chlamydomonadales
C/Date: 02-Mar-2001 #sequence revision 02-Mar-2001 #text change 23-Mar-2001

C/Accession: H86612

Nucleic Acids Res. 28, 2311-2314, 2000

A;Reference number: A86491; MUID:20330349; PMID:10871362

A/Status: preliminary

A;Residues: 1-488 <STO>

A:Experimental source: strain J138

C:Genetics:

A:Gene: htra

C:Superfamily: Helicobacter serine proteinase

Query Match 28.4%; Score 668.5; DB 2; Length 488;
Best Local Similarity 36.1%; Pred. No. 6.9e-36;
Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

QY 10 LAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGSVSMILPDPFQVLVQSEGPAYV 69
DB 10 LAVIVGSSLLALPLLSGQAVG--KKEG---RVSELPOVLLEKISGSGFSKATKATPAVY 63
QY 70 NIQAAPAR--TQNSGNAETSDPLADDPFY-EFFKRL--VNNPEIPQ-SEADDGGL 123
DB 64 YIESFPKQAVTHPSPPRRGRPEYENPF--DYFNDEFNRFPGLPQREKQSKAVR--- 117
QY 124 NFGSGFIISKNGYILINTHVAVGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDATEE 183
DB 118 --GTGFLVSPDGYIVTNHNVVEDTGKIHVTLHDGQKPAFYIGLDPKTDIAVIKI-KSQN 174
QY 184 LPVVKIGPKNLKFGENVAAIGAPFGFDSVTAGIVSAKRS-LPNESYTPPIQTDVAIN 242
DB 175 LPVLSFGNSDHLKVGDMAIAIGNPFGLOATVTVGVISAKGRNQLHIADEFDFIQTDAAIN 234
QY 243 PGNSSGGLPLNLKGVVGINSOIYRSRSGFMGISPAIPIDVAMNVAEQKNTGKVQRGQUG 302
DB 235 PGNSSGGLPLNIDGOVIGVNTAIVSGSGYIGIGFALPSLMANRITDQLIRDQVTRGFVG 294
QY 303 VVIOEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLDGGEIRSSGDLFV 362
DB 295 VTLQPIDAEIAACVLEKVGALVTDVVKSPADKAGLKQEDVITIAVNGKEVDSLMPRN 354
QY 363 MYGAIITPGEKVSIGVWRKGEIITTKAKLGNAAEHTGASSKTDEAPYTEQSGTFPSVESAG 422
DB 355 AVSLMNDTRIVLKVAREGKVEIPVTVSQAPKEDGMS-----ALQRVG 398
QY 423 ITLQTHDSSGKHL-----VVVRVSDAERAGLRHGDILLAV 459
DB 399 IRVQNLTPETAKKLGIAPEYTKGILIIISVEPGSVAASSGIAPQULILAV 446

RESULT 15

G72011
do serine proteinase - Chlamydia pneumoniae (strain CWL029)
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C/Accession: G72011
R.Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
J. Genet. 21, 385-389, 1999
Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A/Reference number: A72000; MUID:99206606; PMID:10192388
A/Accession: G72011
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-488 <ARN>
A/Cross-references: GB:AE001678; GB:AE001363; NID:94377301; PIDN:AAD19116.1; PID:9437730
A/Experimental source: strain CWL029
C:Genetics:
A:Gene: htra
C:Superfamily: Helicobacter serine proteinase

Query Match 28.4%; Score 668.5; DB 2; Length 488;
Best Local Similarity 36.1%; Pred. No. 6.9e-36;
Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

QY 10 LAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGSVSMILPDPFQVLVQSEGPAYV 69
DB 10 LAVIVGSSLLALPLLSGQAVG--KKEG---RVSELPOVLLEKISGSGFSKATKATPAVY 63
QY 70 NIQAAPAR--TQNSGNAETSDPLADDPFY-EFFKRL--VNNPEIPQ-SEADDGGL 123
DB 64 YIESFPKQAVTHPSPPRRGRPEYENPF--DYFNDEFNRFPGLPQREKQSKAVR--- 117

QY 124 NFGSGFIISKNGYILINTHVAVGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDATEE 183
DB 118 --GTGFLVSPDGYIVTNHNVVEDTGKIHVTLHDGQKPAFYIGLDPKTDIAVIKI-KSQN 174
QY 184 LPVVKIGPKNLKFGENVAAIGAPFGFDSVTAGIVSAKRS-LPNESYTPPIQTDVAIN 242
DB 175 LPVLSFGNSDHLKVGDMAIAIGNPFGLOATVTVGVISAKGRNQLHIADEFDFIQTDAAIN 234
QY 243 PGNSSGGLPLNLKGVVGINSOIYRSRSGFMGISPAIPIDVAMNVAEQKNTGKVQRGQUG 302
DB 235 PGNSSGGLPLNIDGOVIGVNTAIVSGSGYIGIGFALPSLMANRITDQLIRDQVTRGFVG 294
QY 303 VVIOEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLDGGEIRSSGDLFV 362
DB 295 VTLQPIDAEIAACVLEKVGALVTDVVKSPADKAGLKQEDVITIAVNGKEVDSLMPRN 354
QY 363 MYGAIITPGEKVSIGVWRKGEIITTKAKLGNAAEHTGASSKTDEAPYTEQSGTFPSVESAG 422
DB 355 AVSLMNDTRIVLKVAREGKVEIPVTVSQAPKEDGMS-----ALQRVG 398
QY 423 ITLQTHDSSGKHL-----VVVRVSDAERAGLRHGDILLAV 459
DB 399 IRVQNLTPETAKKLGIAPEYTKGILIIISVEPGSVAASSGIAPQULILAV 446

Search completed: July 11, 2003, 10:33:59
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:15:32 ; Search time 12 Seconds
(without alignments)
1607.208 Million cell updates/sec

Title: US-09-388-090-4
2353

Perfect score: 1 VFKRYQIFALALCALIAG.....ERAGLRHDEILAVRASPRQ 465

Sequence: 1 VFKRYQIFALALCALIAG.....ERAGLRHDEILAVRASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Match Length	ID	Description
1	747	31.7	504	1	DEGP RHIME
2	731.5	31.1	503	1	DEGP BARHE
3	692.5	29.4	513	1	DEGP BRUBA
4	683	29.0	497	1	DEGP CHLNU
5	683	29.0	497	1	DEGP CHLNU
6	668.5	28.4	488	1	DEGP CHLNU
7	648	27.5	466	1	HMOA HAHIN
8	638	27.1	508	1	DEGP RICCN
9	633.5	26.9	475	1	DEGP SALTY
10	627.5	26.7	455	1	DEGP ECOLI
11	615.5	26.2	513	1	DEGP RICPR
12	610.5	25.9	474	1	DEGP ECOLI
13	594.5	25.3	478	1	DEGP BUCAL
14	571.5	24.3	478	1	DEGP BUCAP
15	502.5	21.4	449	1	HRAA_BACU
16	470	20.0	437	1	DEGI_ARATH
17	459.5	19.5	458	1	YVTA_BACU
18	448	19.0	480	1	HRAI_HUMAN
19	447.5	19.0	355	1	DEGS_ECOLI
20	436.5	18.6	413	1	HRAI_MOUSE
21	435.5	18.5	480	1	HRAI_MOUSE
22	435	18.5	448	1	DEGS_ARATH
23	428	18.2	340	1	DEGS_HAETH
24	414.5	17.6	453	1	HRA2_HUMAN
25	414	17.6	458	1	HRA2_HUMAN
26	412.5	17.5	476	1	HRA4_HUMAN
27	412	17.5	400	1	YXAA_BACU
28	411.3	17.5	458	1	HRA2_MOUSE
29	388	16.5	408	1	HRA3_LACU
30	353	15.0	460	1	HRA3_MOUSE
31	344.5	14.6	321	1	HMOA_ARATH
32	230	9.8	630	1	VABJ_RHIN
33	198.5	8.4	997	1	YNM3_TBAST

34	137.5	5.8	767	1	DLG4_HUMAN	P78352 homo sapien
35	131	5.6	724	1	DLG4_RAT	P31016 rattus norv
36	130	5.5	724	1	DLG4_MOUSE	Q62108 mus musculu
37	128	5.4	2021	1	OMPA_RICCN	Q52657 rickettsia
38	125	5.3	430	1	YKJ5_STRCO	Q9K980 streptomyce
39	118	5.0	280	1	ETA_STAU	P09331 staphylococ
40	115.5	4.9	837	1	APBI_HUMAN	Q02410 homo sapien
41	115	4.9	335	1	YG93_CLOPE	Q8XJ12 clostridium
42	115	4.9	568	1	YELL_DROMD	Q9GP71 drosophila
43	115	4.9	568	1	YELL_DROSU	Q02437 drosophila
44	114.5	4.9	911	1	DLG1_RAT	Q62696 rattus norv
45	114	4.8	904	1	DLG1_HUMAN	Q12959 homo sapien

ALIGNMENTS

RESULT 1
ID DEGP RHIME STANDARD; PRT; 504 AA.
AC Q52894;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP1 OR DEGP OR R01021 OR SMC02365.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=96146524; PubMed=8550509;
RA Glazebrook J., Ichige A., Walker G.C.;
RT Genetic analysis of Rhizobium meliloti bacc-phoA fusion results in
RT identification of degP. Two loci required for symbiosis are closely
RT linked to degP."
RL J. Bacteriol. 178:745-752(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gohrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Maury D.,
RA Pohl T., Portetelle D., Puhler A., Purnelle B., Rampersger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/annotation/>
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: U31512; AAC3669.1; ALU_INIT.
DR EMBL: AF591785; CAC45593.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR000126; Ser protease V8.
DR InterPro; IPR001254; Ser protease V8.
DR Pfam; PF000089; trypsin_1.
DR Pfam; PF00595; PDZ_2.
DR PRINTS; PR00834; PROTEASES2C.

DR PRINTS, PRO00839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 KW Hydrolyase; Serine protease; Periplasmic; Repeat; Signal;
 KM Complete proteome.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 504 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 113 286 CATALYTIC.
 FT DOMAIN 287 378 PDZ 1.
 FT DOMAIN 401 491 PDZ 2.
 FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 244 244 CHARGE RELAY SYSTEM (POTENTIAL).
 FT CONFLICT 14 15
 FT CONFLICT 39 147
 FT PSFANVAVDAYSPAVVSRYVQARERVSDESNTFFPGGRCF
 EDIPEDPLRFRFEAFRENDRADRDNRDRGREGRLRP
 RAQSGFFIRDEGDLTNHNVSDGSA -> AVSMTSTP
 RRRSSPACCHVNASATKATSPISMAAGSRTRKTRCG
 VSSANSLRKMTVPFVGATLAVRKAVYVRGRAPASSSP
 KTVTSPTTSSPTART (IN REF. 1).
 KSADVDLKVNNAKDKGRSALPQIEAQSGRFVALPTQG
 -> NROTTFR (IN REF. 1).
 FT CONFLICT 464 504
 SQ SEQUENCE 504 AA; 5303 MW; D7E82B9981EA23C CRC64;
 Query Match 31.7%; Score 747; DB 1; Length 504;
 Best Local Similarity 36.6%; Pred. No. 5; Le-40;
 Matches 175; Conservative 87; Mismatches 162; Indels 54; Gaps 11;
 10 LMAALCALLA---GCKKAGSFFGADKKEASFVERIETHKDDGSVSMLLPDAQVQSEGP 66
 7 VAGLAVALLTITGIPAEVQASFAEAVRQAAV-----PSFANVADVSP 50
 67 AVVNIOAAPARTQNSGNAETD-----SDPLADSPFEFFKLVNMEIPEOEAD-- 119
 51 AVYSRVQARERVSDESNTFFPGGRCGFEDLPEDHRLRFRFRFAR---ENDRADRW 106
 120 -----DQGLN---FGSGFIISKNGYIITNTHTVAVAGMSIKVLLNDKREYTKLIGSD 168
 107 RDRRGRCGRGLRPRAQSGSFFITDEGYLTNTNHVVSDDSAFIVIMDGTELDAKLVOKD 166
 169 VQSDVALLKIDAEELPVYKIGNPKVKPEBWAIAIGAPGFDNSVTAGVSAKGRSLPN 228
 167 SKTDLVLKVDKRRKTTYVSFADEKRVQDVAVANPNFGLGTGTAGTISAGGRIDGS 226
 229 ESYTFPIQTDVAINPGNSGGLPFLNLKQVGINSOIYSRSGFNGISFALPIDVAMVAE 288
 227 GPYDDVLQYDAANRNGSGGPTNLSEGVGINTAIFSPSGANGVAFALPASAXKVVD 286
 289 QLKNTGKVGQGLGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLS 348
 287 SLIKDGTVSRGWLGVQIOPVTXKIDAEISLGESEANGALVVEPQASPEKKAIGIKGDIYTA 346
 349 LOGGEIRSSGDLPEVMGAIIPGKEVSLGVWRKEEITIKAKLG---NAAEHTGASSKTD 404
 347 LGEPEPKDPRDLARRVAALRPGSTAETVLWRSGSEVTNLEITGLPDAKEPADA---TG 403
 405 EAPYTEQSGSTFVSAGITLQTHTDSGKGLVVVRY---SDAERAGLRHGDIEILAV 459
 404 EAPQPDGQAGEALADLGLTIVTSED--GKGVITIASVDPDSADGR-GLKEGETIVSV 458
 RESULT 2
 DECP BARHE STANDARD; PRT; 503 AA.
 AC P54925;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
 DE (Antigen hcrA).
 CN DECP OR HTRA.
 OS Bartonella henselae (Rochallimaea henselae).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Bartonellaceae; Bartonella.
 OX NCBI_TaxID=36323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Houston-1;
 EX MEDLINE=94299828; PubMed=8027347;
 RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
 RA Goral S., Hager C., Edwards K.;
 RT "Detection of Rochallimaea henselae DNA in specimens from cat scratch
 disease patients by PCR."
 RL J. Clin. Microbiol. 32:942-948(1994).
 CC - SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC - SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL; L20127; AAA97430.1; -.
 DR MEROPS; S01.273; -.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser. protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 2.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 KW Hydrolyase; Serine protease; Periplasmic; Repeat; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-
 FT LIKE.
 FT DOMAIN 286 357 PDZ 1.
 FT DOMAIN 419 466 PDZ 2.
 FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 503 AA; 54114 MW; 6CD9F4743282AF9E CRC64;
 Query Match 31.1%; Score 731.5; DB 1; Length 503;
 Best Local Similarity 39.5%; Pred. No. 4; Le-39;
 Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;
 8 PALAALCALLAGCEKAGSFFGADKKEASFVERIETHKDDGSVSMLLPDAQVQSEGA 67
 15 FSALETLALFPGSC---GSSIMTTTVAHANSV-----FSSLMOQOGFADIVSQVKA 62
 68 VVNIOAAPARTQNSGNAETDSDPLADSDP-----FYEFKLVNMEIPEOEAD 120
 63 VVSQVQSNKKKKKEWFFSDFSTPGFDLPDQHPKKEFQDDYNDKESNKL-QSRHL 121
 121 GGLNFGSGFIISKNGYIITNTHTVAVAGMSIKVLLNDKREYTKLIGSDVQSDVALIKIDA 180
 122 RPIAFSGSFFISSDGYITNTNHVVISDGSYAVVLDGHEINAKKILGTDRITDLAVLVKNE 181
 181 TEBLPVYKIGNPKVKPEBWAIAIGAPGFDNSVTAGVSAKGRSLPNESYTFPIQTDVA 240
 182 KRKFSYVDFGDDSKLRVGDVAVAIQNPFGLTGTGTAGVSAKGRDIGNTVDDFIQIDA 241
 241 INPNSGGLPFLNLKQVGINSOIYSRSGFNGISFALPIDVAMVAEGLKNTGKVGQRO 300
 242 VNRNNGSGPTDLNGKVVGVNTAIFSPSGANGVAFALPAATANEVQGLIKGLVQKDA 301
 301 LGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLSLDGGEIRSSGDL 360
 302 LGVQIOPVTXKIDAEISLGESEANGALITPLKG-PAKAGIKRAGDIYISNGEKINDYRL 360
 361 PVMVGAIIPGKEVSLGVWRKEEITIKAKLGNAEHTGASSXTDEAPYT-EQSGSTFVSVE 419

```

Db      361 AKRIANNSPGEIVTVGVKMSGKEENIKVLTDSMPED--ENMKDGSKYSENGSDEITL 417
Qy      420 SAGITLQHTDSSGKHLVVRV---SDAERAGLRHDEILAV 459
Db      418 DYGLIVAPSDGLG--LVVTVDPDPSDADK-GIRGDIYIVV 457

RESULT 3
DEGP BRUAB          STANDARD;          PRT;          513 AA.
ID  DEGP BRUAB          STANDARD;          PRT;          513 AA.
AC  044597;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Probable serine protease do-like precursor (EC 3.4.21.-).
GN  DEGP OR HTRA.
OS  Brucella abortus.
OC  Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
   Brucellaceae; Brucella.
NCBI_TaxID=235;
[1]
SEQUENCE FROM N.A.
RC  STRAIN=2308;
RX  MEDLINE=9516590; PubMed=7861951;
   Tatum F.M., Chevillie N.F., Morfitt D.;
   "Cloning, characterization and construction of htra and htra-like
   mutants of Brucella abortus and their survival in BALB/c mice.";
   Microb. Pathog. 17:23-36 (1994).
RL  -1- SURCELLULAR LOCATION: Periplasmic (Potential).
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC  -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, U07352; AAA70164.1; -.
DR  InterPro; IPR001478; PDZ.
DR  InterPro; IPR001940; Protease2C.
DR  InterPro; IPR001254; Ser_protease_Try.
DR  Pfam; PF00089; trypsin; 1.
DR  Pfam; PF00595; PDZ; 2.
DR  PRINTS; PR00834; PROTEASES2C.
DR  SMART; SMO0228; PDZ; 2.
DR  PROSITE; PS50106; PDZ; 2.
DR  Hydrolase; Serine protease; Periplasmic; Repeat; Signal.
FT  SIGNAL          1..25
FT  CHAIN           1..513
FT  DOMAIN          125..299
FT  DOMAIN          300..391
FT  DOMAIN          414..500
FT  ACT_SITE        152..152
FT  ACT_SITE        182..182
FT  ACT_SITE        257..257
SQ  SEQUENCE        513 AA; 53483 MW; 53483 MW; DEICEF1959472806 CRC64;

Query Match          29.4%; Score 692.5; DB 1; Length 513;
Best Local Similarity 37.4%; Pred No. 14e-36;
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

Db      9  ALAALCALAGCEKAGSFFGADKKEASFVERIEHTKDDGSVSMLLP---DFAQLVQSRG 65
Qy      14  AAVALLSAL-----AGAFVVTGPGICALNEARAR-----AVHVTPEPPQAGFADLVKVR 61
Db      66  PAVVNIAQAPAPRTONGSGN---AETDSPDLADSDFFYFF-----KRLV 107
Qy      62  PAVVSVRVKQVQETSNRGPOFGFPGFQDLPDGLKRFDFDFGMEPRGDSRSDMRGK 121

```

```

Qy      108  PNMPEIPEEADGGGLNFGSGFIISKNGYIILTNTHVAVAGMSIKVLYLNDKREYAKLIGS 167
Db      122  ANKRPQCHERP-----VAQSGGFVISEDGYVYNNHVVSDGAYTVVLDGTETLDAKLIGA 177
Qy      168  DVQSDVALKTKDTE-ELPVYKIGNPKNLKRGVNAVAGAFPGDNSTAGIVSKGSL 226
Db      178  DPTDLVLKLNPKKRFVYVAFDDNKRVRGDNVAVVAVNGFAGCTVTSIGVSARGDI 237
Qy      227  PNESTYPTFIQDVAINPNSGGPLFNKGOVVGINSOIYSRSGFMGISAIPIDAMNV 286
Db      238  GAGPYDDFIOIDAANVKNNGSGGPAFDLSGEVIGINTALFSPSGSGVGLAFAPISSTACV 297
Qy      287  AEQLKNTGKVQKQQLGYITIEVSYGIAQSPGLDKASGALLAKIIPGSPAERAGLQAGDIV 346
Db      298  VDQLIKKGSVERGWIGVQIOPVTMDIAASTGLAEKGAIVASPODDGPAAKAGIVAGDVI 357
Qy      347  LSLDGGIRSSGGLPVVVGVAITPEKEVSLGWR--KKEETIK-AKLGNAAEHTGASST 403
Db      358  TAVNGETVQDPRLARKVANIAPGEKALTVWRKNKAEINVTJAAMPNDKSGSGSND 417
Qy      404  DEAPYTEQSGTFVSAGITLQHTDSSGKHLVVRV---SDAERAGLRHDEILAV 459
Db      418  NDGQGE-----TLDSGLTVPSED--GKGVVTVDPDPSDADR-GIRSGDIYIV 467

RESULT 4
DEGP CHLMU
ID  DEGP CHLMU          STANDARD;          PRT;          497 AA.
AC  Q9P157;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Probable serine protease do-like precursor (EC 3.4.21.-).
GN  DEGP OR HTRA OR TC0210.
OS  Chlamydia muridarum.
OC  Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX  NCBI_TaxID=83560;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MoPn / N195;
RX  MEDLINE=20150255; PubMed=10684935;
   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
   RA  White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
   RA  Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
   RA  Gilm M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
   RA  Eisen J., Fraser C.M.;
   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
   RT  pneumoniae AR39.";
   Nucleic Acids Res. 28:1397-1406 (2000).
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC  -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE002289; AAF39082.1; -.
DR  MEROPS; S01.273; -.
DR  TIGR; TC0210; -.
DR  InterPro; IPR001478; PDZ.
DR  InterPro; IPR001940; Protease2C.
DR  InterPro; IPR000126; Ser_protease_V8.
DR  InterPro; IPR001254; Ser_protease_Try.
DR  Pfam; PF00089; trypsin; 1.
DR  Pfam; PF00595; PDZ; 2.
DR  PRINTS; PR00834; PROTEASES2C.
DR  PRINTS; PR00839; VBPROTEASE.
DR  SMART; SMO0228; PDZ; 2.
DR  PROSITE; PS50106; PDZ; 2.

```

```

KW Hydrolyase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 497
FT DOMAIN 128 289
FT DOMAIN 290 381
FT DOMAIN 394 485
FT ACT SITE 143 143
FT ACT SITE 173 173
FT ACT SITE 247 247
SQ SEQUENCE 497 AA; 53294 MW; 8765F350AC68B8F CRC64;

Query Match
Best Local Similarity 37.4%; Score 683; DB 1; Length 497;
Matches 173; Conservative 77; Mismatches 154; Indels 58; Gaps 11;

OY 28 FGADKKEASFERIENTKDDGSVSM--LLPD----FAQLVSGEPVAVNIOAAPARTQ- 80
DB 22 YSAPKRDSSGICLAASQSDRELSDLEKVSRRGKVAQAATPGVVTIENPKTGNOA 81
OY 81 -----NGSGNAETDSDPLADSDPFYEFKRL--VNNPEIPOEADDDGGLNFGSGFIISK 133
DB 82 IASPKRGKGFENPDYFND-----EFNRPFGLPESHREGRRPQORAVR--GTGFIYSE 134
OY 134 NGYILTNTHVAVMGSIKVLNDRREYTAKLIGSDVQSDVALLKIDATEELPVVKIGNPK 193
DB 135 DGVVVTHNHVVEDAGKIHVTLHDGQKYAKIIGIDPKTDLAIVIKQA-KNLPFLTFGNSD 193
OY 194 NLKPGEMVAALGAPFGFNSVTAGIVSAKGRS-LPNESYTPFIOTDVAINNGSGGLFN 252
DB 194 QLOIGDMSTALGNPFGLOATVTGVIISAKGRNQLHIVDFEDFIOTDAINNGSGGLFN 253
OY 253 LKQGVGINSQIYSRSGFGMGISFAIPIDVAMNVAEQLKNTGKYQSGQLGVIIOEVSYGL 312
DB 254 INGQVIGNTVAIVSGSGGVIIGIIPSLMAKRVTDQLISDQVTRGFLGVTLLQPIDSEL 313
OY 313 AQSFGIDKASGALIAKILPGSPAPRAGLQAGDIVLSLDGGEIRSSGDLPMVNGAITPGKE 372
DB 314 AACYKLEKYVALITVYVKGSPAEKAGLRQEDVIVAANGKEVESLSALRNALSLMMPGTR 373
OY 373 VSLGVARKKEEITIKAKLGNAAEHTGASSKTDEAPYREQ---SGTFVSASAITIQTQ 428
DB 374 VVLKVRREGKFI-----ELPVTVTQIPAEDEVSAIQKMGVRYQNL 413
OY 429 T-----DSSGKHLVVRVSDAABERAGLRHGDDEILAV 459
DB 414 TPEICKKGLASDPTRGIFVSVSEAGSPRAAGVIRGQILILAV 455

DUT 5
DEGP CHLTR STANDARD; PRT; 497 AA.
AC P18584; O84830.
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
GN immunogenic protein) (SK59).
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=813;
(1)
RN SEQUENCE FROM N.A.
RP STRAIN=Sevovar L2;
RX MEDLINE=90337348; PubMed=2379836;
RT Kahane S., Weinstein Y., Sarov I.;
RT "Cloning, characterization and sequence of a novel 59-kDa protein of
RT Chlamydia trachomatis.";
RL Gene 90:61-67(1990).
RC SEQUENCE FROM N.A.
RP STRAIN=D/UW-3/Cx;
RX MEDLINE=9000809; PubMed=9784136;

```

```

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RT Science 282:754-759(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -1- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 KDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AE001355; AAC68420.1; -.
DB EMBL; M31119; AAA23116.1; -.
DB PDB; 1ZDAGE; P18584; -.
DB InterPro; IPR001478; PDZ.
DB InterPro; IPR001940; Protease2C.
DB InterPro; IPR000126; Ser_protease_V8.
DB InterPro; IPR001254; Ser_protease_Try.
DB Pfam; PF00089; trypsin; 1.
DB Pfam; PF00595; PDZ; 2.
DB PRINTS; PR00834; PROTEASE2C.
DB PRINTS; PR00839; V8PROTEASE.
DB SMART; SM00228; PDZ; 2.
DB PROSITE; PS0106; PDZ; 2.
KM Hydrolyase; Serine protease; Repeat; Signal; Antigen;
KW Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 497
FT DOMAIN 128 289
FT DOMAIN 290 381
FT DOMAIN 394 485
FT ACT SITE 143 143
FT ACT SITE 173 173
FT ACT SITE 247 247
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31B884A38BA CRC64;

Query Match
Best Local Similarity 29.0%; Score 683; DB 1; Length 497;
Matches 171; Conservative 75; Mismatches 158; Indels 58; Gaps 11;

OY 28 FGADKKEASFERIENTKDDGSVSM--LLPD----FAQLVSGEPVAVNIOAAPARTQ- 80
DB 22 YSAPKRDSSGICLAASQSDRELSDLEKVSRRGKVAQAATPGVVTIENPKTGNOA 81
OY 81 -----NGSGNAETDSDPLADSDPFYEFKRL--VNNPEIPOEADDDGGLNFGSGFIISK 133
DB 82 IASPKRGKGFENPDYFND-----EFNRPFGLPESHREGRRPQORAVR--GTGFIYSE 134
OY 134 NGYILTNTHVAVMGSIKVLNDRREYTAKLIGSDVQSDVALLKIDATEELPVVKIGNPK 193
DB 135 DGVVVTHNHVVEDAGKIHVTLHDGQKYAKIIGIDPKTDLAIVIKQA-EKLPLTFGNSD 193
OY 194 NLKPGEMVAALGAPFGFNSVTAGIVSAKGRS-LPNESYTPFIOTDVAINNGSGGLFN 252
DB 194 QLOIGDMSTALGNPFGLOATVTGVIISAKGRNQLHIVDFEDFIOTDAINPNSSGGLFN 253
OY 253 LKQGVGINSQIYSRSGFGMGISFAIPIDVAMNVAEQLKNTGKYQSGQLGVIIOEVSYGL 312
DB 254 INGQVIGNTVAIVSGSGGVIIGIIPSLMAKRVTDQLISDQVTRGFLGVTLLQPIDSEL 313
OY 313 AQSFGIDKASGALIAKILPGSPAPRAGLQAGDIVLSLDGGEIRSSGDLPMVNGAITPGKE 372
DB 314 ATCYKLEKYVALITVYVKGSPAEKAGLRQEDVIVAANGKEVESLSALRNALSLMMPGTR 373

```

QY 373 VSLGWRKKEEITITKAKLGNAAEHTGASSKTDEAPYTEQ-----SGTFVSASGITLQTH 428
 Db 374 VLKIVREG-----KTIETPTVTVQIPREDGVSALQKGVQRQNI 413
 QY 429 T-----DSSGKHLVYVRVSDAERAGLRHGDILLAV 459
 Db 414 TPEICKKGLADTRIGLIVVAVEAGSPAAAGVAPQQLILAV 455

RESULT 6
 DEGP_CHLPN STANDARD; PRT; 488 AA.
 AC Q926T0; Q9JQD7; Q9K1W4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Probable serine protease do-like precursor (EC 3.4.21.-).
 OS DEGP OR HTRA OR CPN0979 OR CP0877.
 NC Chlamydia pneumoniae (Chlamydia pneumoniae).
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_TaxID=83558;
 [1]
 [2]
 [3]
 [4]
 [5]
 [6]
 [7]
 [8]
 [9]
 [10]
 [11]
 [12]
 [13]
 [14]
 [15]
 [16]
 [17]
 [18]
 [19]
 [20]
 [21]
 [22]
 [23]
 [24]
 [25]
 [26]
 [27]
 [28]
 [29]
 [30]
 [31]
 [32]
 [33]
 [34]
 [35]
 [36]
 [37]
 [38]
 [39]
 [40]
 [41]
 [42]
 [43]
 [44]
 [45]
 [46]
 [47]
 [48]
 [49]
 [50]
 [51]
 [52]
 [53]
 [54]
 [55]
 [56]
 [57]
 [58]
 [59]
 [60]
 [61]
 [62]
 [63]
 [64]
 [65]
 [66]
 [67]
 [68]
 [69]
 [70]
 [71]
 [72]
 [73]
 [74]
 [75]
 [76]
 [77]
 [78]
 [79]
 [80]
 [81]
 [82]
 [83]
 [84]
 [85]
 [86]
 [87]
 [88]
 [89]
 [90]
 [91]
 [92]
 [93]
 [94]
 [95]
 [96]
 [97]
 [98]
 [99]
 [100]

DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 KW Hydrolyase; Serine protease; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 488
 FT DOMAIN 119 280
 FT DOMAIN 281 372
 FT DOMAIN 388 476
 FT ACT_SITE 134 134
 FT ACT_SITE 164 164
 FT ACT_SITE 238 238
 SQ SEQUENCE 488 AA; 52311 MW; 0BE7E0F88F106F49 CRC64;

Query Match 28.4%; Score 668.5; DB 1; Length 488;
 Best Local Similarity 36.1%; Pred. No. 4.3e-35;
 Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

QY 10 LAAALCAALLAGCEKXGFFGADKKEASFVERIEHTKDDGSVSMLLPDFAQIVQSGRPAYV 69
 Db 10 LAVLVGSSLLALPLSGQAVG--KKE-----RVSELQVVLKKEISGFSKVATKATPAV 63
 QY 70 NIQAAPAR--TQNGSNAETSDPLADSPFY-EFFKRL--VPMPEIPQ-EEADGGL 123
 Db 64 YIESFPKSOAVTHPSPGRRGFYENPF---DYFNDEFFNRPGLPSQREKPKSKEAVR--- 117
 QY 124 NFGSGFIISKNGYLTITTHVVAAGMSIKVLNDRKYTAKLIGSDVQSPVALTKIDATEE 183
 Db 118 --GTGFLVSPGXYVTNNHVEDTGKIHVLHDGQKPAIVYGLDPKIDLAVIKI-KSON 174
 QY 184 LPVVKINPKNLKGEVVAALGAPFGEDNSVTAGIVSAKGRS-LPNESTPTPTQDVAIN 242
 Db 175 LPVLSFNSDHLKAKGDVAIALGNFGLQATVTGVIASAKRQHLIADPEDTQIDALIN 234
 QY 243 PGNNGGFLPNKGVQVINSQIYRSRSGFMGISPAIPIDVAMNVAOLKNTKVKVQGLG 302
 Db 235 PGNNGGFLPNIDGGVIGVNTAIVSGSGGYIGIFAIPEIMANRIIDQIRDOQVRGFLG 294
 QY 303 VTIQEVSYGLAQSGFLDKASGALAKILPGSPABAGIQAQGIYVSLGGGEIRSSGDLV 362
 Db 295 VTIQPIDELAACKLEKRYVALTVDVKSPADGAKQGEVITAYVNGKEDVLSMPFN 354
 QY 363 MVGAITPEKEVSLGWRKKEEITITKAKLGNAAEHTGASSKTDEAPYTEQSGTFVSASG 422
 Db 355 AVSLMNPDTRIVLKVAEBGKYLEIPVYVSOAPKEGNS-----ALQRVG 398
 QY 423 ITLQTHDSSGKHL-----VVVRVSDAERAGLRHGDILLAV 459
 Db 399 IRVQNLTPETAKKLGIAPETKGIILISVEPSVAASGIAPQQLILAV 446

RESULT 7
 HTOA_HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Probable periplasmic serine protease do/hda-like precursor (EC 3.4.21.-).
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 NCBI_TaxID=727;
 [1]
 [2]
 [3]
 [4]
 [5]
 [6]
 [7]
 [8]
 [9]
 [10]
 [11]
 [12]
 [13]
 [14]
 [15]
 [16]
 [17]
 [18]
 [19]
 [20]
 [21]
 [22]
 [23]
 [24]
 [25]
 [26]
 [27]
 [28]
 [29]
 [30]
 [31]
 [32]
 [33]
 [34]
 [35]
 [36]
 [37]
 [38]
 [39]
 [40]
 [41]
 [42]
 [43]
 [44]
 [45]
 [46]
 [47]
 [48]
 [49]
 [50]
 [51]
 [52]
 [53]
 [54]
 [55]
 [56]
 [57]
 [58]
 [59]
 [60]
 [61]
 [62]
 [63]
 [64]
 [65]
 [66]
 [67]
 [68]
 [69]
 [70]
 [71]
 [72]
 [73]
 [74]
 [75]
 [76]
 [77]
 [78]
 [79]
 [80]
 [81]
 [82]
 [83]
 [84]
 [85]
 [86]
 [87]
 [88]
 [89]
 [90]
 [91]
 [92]
 [93]
 [94]
 [95]
 [96]
 [97]
 [98]
 [99]
 [100]

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utechtback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritschman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.",
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E. COLI HTRA
 CC (PROTEASE DO) AND HHOA.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32805; AAC22906.1; -.
 CC MEROPS: S01.274; -.
 DR TIGR: H11259; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00595; PDZ; 2.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 DR KEGG: Hydroxylase; Serine protease; Periplasmic; Repeat; Signal;
 DR Complete proteome.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 466 PROBABLE PERIPLASMIC SERINE PROTEASE
 FT DO/HOA-LIKE.
 FT DOMAIN 270 361 PDZ 1.
 FT ACT_SITE 120 120 PDZ 2.
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 466 AA; 49434 MW; ED050A00047B5851 CRC64;

Query Match 27.5%; Score 648; DB 1; Length 466;
 at Local Similarity 37.6%; Pred. No. 8e-34;
 cche 166; Conservative 76; Mismatches 163; Indels 36; Gaps 13;

QY 29 GADKEASFEVERIEHTKDDGSVSMLLPDPFAQLVQSEGPVAVNIQAPAPRTONGSNAET 88
 DB 14 GLSVLSTSVAVH-----AQATLPSFVSEQNSLAPMLEKQ--PAVVTLSVEGKAKV 63
 QY 89 DS-DPLADSDP--FVFF-KRLVPMPEIQEADDDGLNFGSGFTI-SKNGYLLTTHV 143
 DB 64 DSRSPFLDIDPEEFKFFPGDFPAEFGGSGESKRNFGTL--GSGVIINASKGYLLTNHV 121
 QY 144 VAGMGSIKVLNDKREYTKALIGSDVQSDVALKIDATEELPVYKIGNPKMLKGEVAA 203
 DB 122 IGGADKITVQLODGRFKAKLVGKDESDIALVQLEKPSNLTEIKFADSDKLKRGDTTVA 181
 QY 204 IGAPFGDNSVTAGIVSANGRSILPNES--YTPFIQTDVAINPNSGGPLFNLKQVAVIN 261
 DB 182 IGNPFLGQTVTSIGVIALGRSTGSDGYENTYIQTDAAVNRNGSGALVNLINGELGIN 241
 QY 262 SGIYRSGSGFMGISAIPIDIVANVAEQKNTGKVGQGLGVIIQEVSYGLAOSFGGLDKA 321
 DB 242 TAILSPSGGNAGIAFAIPSNQASNLVQQLIEFGQVRRGLIGIKGELINADLAKAFNVAQ 301
 QY 322 SGALAKLILGSPAPBAGLOAGDIVLSLOGEIRSSGDLPVMGATIRPGKSVSGVWRKG 381
 DB 302 OGAFVSEVLPKSAERKAGLADGITRANNGOKISSFAIRAKIATTGAGKEISLTYLRDG 361

QY 382 EETIKAKLGNAAHRTGASSTIDEAPYTEQSGTFESVSGAGITLQTHDSKGLVAVR 441
 DB 362 KSHIVKML-QADSSQSSKXT-ELPALD-----GATLKDYDAKVGKGIETKX 408
 QY 442 ---SDAERAGLRHGDDELAV 459
 DB 409 QPNSLAQR-GLKSGDITIGI 428

RESULT 8

DEGP_RICCN

ID DEGP_RICCN STANDARD; PRT; 508 AA.

AC Q92JAL;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable serine protease do-like precursor (EC 3.4.21.-).

GN DEGP OR HTRA OR RC0166.

OS Rickettsia conorii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=781;

RN [1] SEQUENCE FROM N.A.

RP STRAIN=Malish 7;

RX MEDLINE=2142074; PubMed=11557893;

RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,

RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,

RA Raoult D.

RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";

RL Science 293:2093-2098(2001).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.

CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE008583; AL020704.1; ALT_INIT.
 CC InterPro: IPR001478; PDZ.
 CC InterPro: IPR001940; Protease2C.
 CC InterPro: IPR001254; Ser_protease_Try.
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00834; PROTEASES2C.
 CC SMART: SM00228; PDZ; 2.
 CC PROSITE: PS50106; PDZ; 1.
 CC KEGG: Hydroxylase; Serine protease; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 508 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 119 284 PDZ 1.
 FT DOMAIN 286 377 PDZ 2.
 FT DOMAIN 413 497 PDZ 2.
 FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 164 164 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 508 AA; 55599 MW; D2F53A690ECDAD7 CRC64;

Query Match 27.1%; Score 638; DB 1; Length 508;
 Best Local Similarity 36.8%; Pred. No. 3.8e-33;
 Matches 153; Conservative 85; Mismatches 152; Indels 26; Gaps 11;

QY 57 PAQLVQSEGPVAVNIQAPAPRTONGSNAETSDPLADSPFYEFKRLVNMPEIQE 116
 DB 56 FADIVERLIPAVNVISTREYVNSK--SENAE--KDPL--QEKVNDFLKLT--NIP-LNLE 106
 QY 117 EADD--GGLNFGSGFTISKNGYILLNTHVAVAGMGSIKVLNDKREYTKALIGSDVQSDVA 174
 DB 107 EVDQTPKSPVLSGSGFTIEPNGLIVTNVIANVDKINIKLADNTELSAKLIGNDKTDLA 166


```

QY 175 LKIDATEELPVVKIGNPNKLRGKGVAAIGAPFG-FDNSVTAGIVSAKGRSLPNES---230
D 167 LKIDSEELPFVEBGRDSDNARVGDWVIAIGNFGMLGGVTGIISSKRDIDIDPTDN1 226
QY 231 YTPFIQTVAINPNSGGPLFNLKQGVVNGINSQIYRSGSGFMISPAIPIDVANNVAEQ1 230
D 227 VDNFIQTAAIINNNGSGGFMFNLDOFVIGVNTAIFSPPLGNICIGFPAIPNTAKPIIERL 286
QY 291 KNTGKVGKRGOLGVIIOEVSYGLAQSFGDLKASGALAKLIPGSPARAGAOADIVLSLD 350
D 287 KKDGVNSGRGLVTTIQDITFEDISEGIGLKNTKRGVLAQKQEDGPGDAKAGTKTDIIEFA 346
QY 351 GGEIRSGDLPVWGAITPKEVSLGVWRKGEIITTKAKLGNAAEHT--GASSTDEAPY 408
D 347 DIPVNTKTKLRVLIADAPIDQEVKVKILRDKKELEPIKITSNDEEVTYKOSTETNKKEI 406
QY 409 TEQSGTGSVEAGITL-----QTHDSSGKHLVVVRVSPDAERAGLRHDEI 456
D 407 TNKENNMLSTKNNTTFGNLTTELKROKRTIPODKMGIVITNID-EESSSPKIDLI 461
DEGP_SALTY STANDARD; PRT; 475 AA.
ID DEGP_SALTY STANDARD; PRT; 475 AA.
AC P26982;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protease deo precursor (EC 3.4.21.-).
OS Salmoneilla typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91251770; PubMed=1645840;
RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
RA Ali T., Miller I., Hormaeche C.,
RA "The role of a stress-response protein in Salmoneilla typhimurium
RA virulence."
RA Mol. Microbiol. 5:401-407(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=L72 / SGGC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RC McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RC Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RC Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RC Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RC Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
RA LT2."
RA Nature 413:852-856(2001).
RL -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
CC SPECIFICITY WITH HHOA/DEGO.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54548; CAA38420.1; -.

```

```

DR EMBL; AE008704; AAL19173.1; -.
DR PIR; S15337; S15337.
DR PIR; S21327; S21327.
DR MEROPS; S01.273; -.
DR Strygene; SG10173; degP.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser protease_Try.
DR Pfam; PF00089; trypsin_1.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydroxylase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 475 PROTEASE DO.
FT DOMAIN 281 372 PDZ 1.
FT DOMAIN 378 467 PDZ 2.
FT ACT_SITE 132 132 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 162 162 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 475 AA; 49315 MW; 86B65BF3C1A289F CRC64;

Query Match 26.9%; Score 633.5; DB 1; Length 475;
Best Local Similarity 34.1%; Pred. No. 6.7e-33;
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;

QY 44 TKDDGSVSMILPDAQVQSEGPVAV--NIOAAPARTONGSNAE---TDSPLASD 97
D 29 TSSSAMTAQOMPSTAPMLKMPVSVSINVGSSTVTPRPMRPNQOQFGDSPFCODS 88
QY 98 PVEFFKRLVNNMEIQEADDG-----INFGSGFTII-SKNGYILTTHVYAGWG 148
D 89 P-----QNSPFC--QGGGANGGQOQKFMALGSGVLIIDAKKGVVNNHVDNAS 138
QY 149 SIKVLNDKREYTKALIGSDVQSDVALLKIDATEELPVVKIGNPNKLRGKGVAAIGAPF 208
D 139 VIKVQLSDGKRFKAVKVGKQPRSDIALIQNPKNLTAKLADSDALNVGDVTAIGNF 198
QY 209 GPDNSVTAGIVSAKGRS-LPNESTYPTPIQTDVAINPNSGGPLFNLKQGVVNGINSQIYS 267
D 199 GLGETVTSIGVISAIGRSGLVNENYENFIQTDAAINRNGSGALVNLNGLIGINTAILAP 258
QY 268 SGFRMGISFALPIDVANNVAEQLNKTKGVQSGQGVIIQEVSYGLAQSFGDLKASGALIA 327
D 259 DGGNIGIGFALIPSNMWNLTQWVEYGVQVKKGEIGMETELNSLAKMKVDAQRCAPVS 318
QY 328 KILPGSPABRAGLQAGDIVLSLDGEIRSGDLPVWGAITPKEVSLGVWRKGEIITTK 387
D 319 QWPNSSPAKRGITAGVITSLNKPITSSPAAALAAQVGTMPVGSKISGLREGAITYN 378
QY 388 AKLGNAAEHTGASSKTDAPYTEQSGTFYSVA-----GITLQTHDSSGKHLVVVR 440
D 379 LEL-----QSSQSQYDSSSTIFSGIGAEKSMNKQDGVVSVK 418
QY 441 VSDAERAGLRHDEILAVRASP 463
D 419 ANSPAQIGLKKDPIVIGANQP 441

RESULT 10
ID DEGO_ECOLI STANDARD; PRT; 455 AA.
AC P39099;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protease dego precursor (EC 3.4.21.-).
GN DEGO OR HHOA OR B3234.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

```

OK NCBI_TaxID=562;
 RP (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Base S., Gu O., Goddard A.;
 RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
 RP (2)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RL "Characterization of degQ and degS, Escherichia coli genes encoding homologs of the DegP protease.";
 J. Bacteriol. 178:1146-1153(1996).
 RP (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
 RL "The complete genome sequence of Escherichia coli K-12.";
 Science 277:1453-1474(1997).
 CC -1- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U15661; AAC3992.1; -
 DR EMBL; U32495; AAC44005.1; -
 DR EMBL; U18997; AAA58036.1; -
 DR EMBL; AB000402; AAC76266.1; -
 DR MEROPS; S01.274; -
 DR SWISS-2DPAGE; P39099; COLI.
 DR Ecogene; EG12612; degQ.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 2.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
 KM Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 455 PROTEASE DEGP.
 FT DOMAIN 258 349 PDZ 1.
 FT ACT_SITE 355 447 PDZ 2.
 FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 139 139 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;
 QY Query Match 26.7%; Score 627.5; DB 1; Length 455;
 DB Best Local Similarity 37.2%; Pred. No. 1,5e-32;
 DB Matches 155; Conservative 72; Mismatches 141; Indels 49; Gaps 7;
 QY 54 LPDFALVQSESPAVVNIQAPAPRTQNGSGNAETDSDPLADSDPFYEFKRLVPMNPEI 113
 DB 39 LPSLAPMLEKVLPAVSVAV-----EGTASQGGKIRFEFKFKFGDDLDPDPAQ 86
 QY 114 POEADDDGGLNGSGFII-SKNGYIITNTHVAVGMSIKVLLDKRKYAKLIGSDVQSD 172
 DB 114 POEADDDGGLNGSGFII-SKNGYIITNTHVAVGMSIKVLLDKRKYAKLIGSDVQSD 172

DB 87 PFE-----GLSGGVIIINASKGYVLTNNHVINQAKISIQLDNGREDPAKLIGSDQSD 139
 QY 173 VALKIDATEELPVYKIGNPKLKGEMWAAIGAFPGFDSNTAGIVAKGRS-LPNESY 231
 DB 140 IALLQIQNPSTLTQIALNDSKLRVAVAGNPFGLGQRTSITVSALGSSGLNLEL 199
 QY 232 TPTIQTVAINPNSGGPFLNLKQGVGINSQISRSQGMGISPAIPDVANVAEQK 291
 DB 200 ENFIQTDAISIRNGSGALLNLNGLNELIGINTAILAPGGSGVIGIPSPNMARTLAQOLI 259
 QY 292 MTGVQVQGGQGVIIQEVSYGLAQSFGDKAGALIAKLTPGSPARAGLQANDIYLSLG 351
 DB 260 DFGIKRGLGIGIKTESADIAKAFNLDVQGFVSEVLPGSGSAKAGVADIIITSLLG 319
 QY 352 GEIRSGDLPVMVGAITPGKEVSLGVMKGEERTIKAKLGNABHTGASSTKDEAPYREQ 411
 DB 320 KPLNSPFLERLRIRITTEGTGVYKGLRNGRPLEVEVLDTST----- 362
 QY 412 QSGTFVESAGITLQHTDSSG-----KHLVVRVSDAERAGLRHGDILLAV 459
 DB 363 -SSSASAEMITPALEGATLSDQKDGKGIGIKIDEVKGSPAAQ-AGLQKDVIIIGV 417
 RESULT 11
 DEGP_RICPR
 ID DEGP_RICPR STANDARD; PRT; 513 AA.
 AC 005942;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable serine protease do-1ike precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR RP124.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsia.
 CX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O., Sichenitz-Ponten T., Aismark U.C.M., Podowski R.M., Naeslund A.K., Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RL "The genome sequence of Rickettsia prowazekii and the origin of mitochondria.";
 RT Nature 396:133-140(1998).
 RL Nature 396:133-140(1998).
 RN [2]
 RP SEQUENCE OF 161-513 FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=97419517; PubMed=9274032;
 RA Andersson J.O., Andersson S.G.E.;
 RT "Genomic rearrangements during evolution of the obligate intracellular parasite Rickettsia prowazekii as inferred from an analysis of 52015 bp nucleotide sequence.";
 RL Microbiology 143:2783-2795(1997).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ235270; CAA14593.1; -
 DR EMBL; Y11782; CAA72471.1; -
 DR MEROPS; S01.273; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.

[1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Devine K.M.;
 RT "Sequence of the *Bacillus subtilis* genome between xlyA and ykoR."
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Azevedo V., Bortner M.G., Beesies P., Bortner A., Borchert S., Bion S., Bortner R., Bortner L., Brans A., Braun M., Brignell S.C., Bron S., Brouillette S., Brusch C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codoni J.J., Conerton I.F., Cummings N.J., Daniel R.A., Dentzer F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabre C., Ferrari E., Follmer D., Fritz C., Fujita M., Fujita Y., Funa S., Gallazzi A., Gallardo N., Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karmata D., Kasahara Y., Klaener-Blanchard M., Klein C., Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapine A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nohack M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Paro V., Pohl T.M., Portelle D., Portollik S., Prescott A.M., Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadle Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi Y., Sekowska A., Seror S.J., Serro P., Shin B.S., Solde B., Sorokin A., Taccioni E., Takagi T., Takehashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A., Toesto V., Uchiyama S., Vandenbol M., Vamier F., Vaasatoti A., Viati A., Wambutt R., Wedler E., Wedler H., Weitzengger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yamamoto K., Yata K., Yoshida K., Yoshikawa H., Zumbstein E., Yoshikawa H., Zumbstein A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*."
 RL Nature 390:249-256(1997).
 RN [3]
 RP NATURE 390:249-256(1997).
 RC TRANSCRIPTIONAL REGULATION.
 RX STRAIN=168;
 RA PubMed=10692364;
 RT "Expression of ykda, encoding a *Bacillus subtilis* homologue of HtrA, is heat shock inducible and negatively autoregulated."
 RL J. Bacteriol. 182:1592-1599(2000).
 RN [4]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX PubMed=11133960;
 RA Noone D., Howell A., Collety R., Devine K.M.;
 RT "ykda and ykta, HtrA-like serine proteases in *Bacillus subtilis*, engage in negative autoregulation and reciprocal cross-regulation of ykda and ykta gene expression."
 RL J. Bacteriol. 183:654-663(2001).
 RN [5]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX PubMed=11555295;
 RA Hyvrylaeinen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P., Vilkainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M., Kontinen V.P.;
 RT "A novel two-component regulatory system in *Bacillus subtilis* for the survival of severe secretion stress."
 RL Mol. Microbiol. 41:1159-1172(2001).
 CC -1- FUNCTION: May be involved in processing, maturation, or secretion of extracellular enzymes.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -1- INDUCTION: Transcription is cases dependent. Induced by heat shock during exponential growth and by heterologous amyloses at the

CC transition phase of the growth cycle. Negatively regulates its own
 CC expression during exponential growth and during heat shock.
 CC -1- MISCELLANEOUS: Inactivation results in compensating overexpression
 CC of ykta, especially during stress conditions.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AJ002571; CA05570.1; -
 DR EMBL: Z99110; CAB1347.1; -
 DR MEROPS: S01.273; -
 DR Subtilisin; B012608; htrA.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser. protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS0106; PDZ; 1.
 KW Hydroxylase; Protease; Serine protease; Heat shock; Transmembrane;
 KW Complete proteome.
 FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 45 67 POTENTIAL.
 FT DOMAIN 68 449 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 103 108 POLY-SER.
 FT DOMAIN 146 152 POLY-SER.
 FT DOMAIN 348 437 PDZ.
 FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 449 AA; 47713 MW; E12B07A9018EE414 CRC64;

Query Match 21.4%; Score 502.5; DB 1; Length 449;
 Best Local Similarity 35.5%; Pred. No. 1.1e-24;
 Matches 138; Conservative 57; Mismatches 135; Indels 59; Gaps 11;

QY 30 ADKKEASPERIEHTKDDGVSMILPFAQLVQSEGPVWVNGAAPARTONGSNAED 89
 DB 98 SEKKSSSSSAFSEBSKISDMVEDS-----PAIVGTTNLAQGSNSLFGSSSD 150
 QY 90 SDPLADSDPFYEFKRLVPMNPEIQEADGGLNFGSGFIISK-NG--YILTNTHVAG 146
 DB 151 S-----SEDTESGS---GSGVIFKENGKXAYIITNNHVVEG 183
 QY 147 MGSIKVLNDRREYTKLIGSDVQSLKIDATEELPVYKIGPKXUKPEEWAATGA 206
 DB 184 ASSIKVSLYDTEYATKAVGSDSLDLAVLIQISDHWTKVNAFGSSDLRTBEYAIAD 243
 QY 207 PFGFD--NSVTAGYVSAKRSI-----PNESTYPIQTDVAINPNPNSGGPFLNLKGQV 259
 DB 244 PLGKDLRTVQGIIVSGVDRVSMSTSGESTSINVITDAALPNPNSGGPFLNTDQKIVG 303
 QY 260 INSQIYRSRGFMGICISFAIPIDVAMNVANQKNTGKVGQGLVIT-----QEV 309
 DB 304 INSKKIBED-DVEIGFAPISNDVXKPIABELLSKQIRPYIGVGMLEQVPQVQEST 362
 QY 310 YGLAQSRLDRAAGALAKIIPGSPAPAGIQAGDIVSLDGGEIRSSGDL-PVNYGAIT 368
 DB 363 LGL---FGSQLNKVVYIREVASGSPAEKAGLKAEDIIILGKKEIDTSGELNNIYKDAK 419
 QY 369 PGKEVSLGVWKKSGEITKATLGNAAEHT 397
 DB 420 IGTVEVKILNKGKEMTKIKLDOKEEKT 448

Search completed: July 11, 2003, 10:31:39
Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:25:17 ; Search time 83 Seconds
(without alignments)
1154.360 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKEYQVFAALCAALAG.....BRAGLRHGDDEILAVRASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Maximum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriap:*
17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2246	95.5	499	16	Q9JVT1
2	989.5	42.1	474	2	Q9ALU1
3	985.5	41.9	474	16	Q57155
4	900	38.2	490	16	Q8XPT5
5	896.5	38.1	505	16	Q8Y016
6	894.5	38.0	481	2	Q9AOD1
7	884	37.6	511	2	Q8RTK2
8	862.5	36.7	514	2	Q9PBA3
9	746	31.7	513	16	Q9BCS8
10	743	31.6	473	2	Q44476
11	735	30.6	523	16	Q8UG08
12	719.5	30.6	514	16	Q985F9
13	716	30.4	514	16	Q8UDS7
14	693.5	29.5	513	16	Q8Y332
15	683	29.0	524	16	Q8Y132
16	673	28.6	504	16	Q98KJ1

17	667.5	28.4	403	16	Q8XV99	Q8XV99 ralsionia s
18	664.5	28.2	500	2	Q44652	Q44652 bruceella ab
19	660	28.0	453	2	Q9FD11	Q9FD11 aeromonas h
20	653.5	27.8	457	16	Q8ZB58	Q8ZB58 yersinia pe
21	649	27.6	481	16	Q8ZB56	Q8ZB56 yersinia pe
22	639.5	27.2	478	2	P74978	P74978 yersinia en
23	639	27.2	474	16	Q8YHL4	Q8YHL4 bruceella me
24	637	27.1	437	2	Q68197	Q68197 haemophilus
25	637	27.1	459	16	Q9CMS7	Q9CMS7 pasteurella
26	637	27.1	474	2	Q44596	Q44596 bruceella ab
27	636.5	27.1	475	16	Q8Z9B0	Q8Z9B0 salmonella
28	636	27.0	503	16	Q9Z6C8	Q9Z6C8 rhizobium m
29	636	27.0	530	16	Q9A4S2	Q9A4S2 caulobacter
30	635.5	27.0	477	2	Q8RSJ1	Q8RSJ1 klebsiella
31	632.5	26.9	456	16	Q9KJF5	Q9KJF5 vibrio chol
32	632	26.9	463	2	Q68198	Q68198 haemophilus
33	631	26.8	429	16	Q8YVZ0	Q8YVZ0 anabaena sp
34	623.5	26.5	455	16	Q8X9F1	Q8X9F1 escherichia
35	621.5	26.4	500	2	Q9KJN6	Q9KJN6 myxococcus
36	617.5	26.2	455	16	Q8Z3E6	Q8Z3E6 salmonella
37	614.5	26.1	459	16	Q9WZ41	Q9WZ41 thermotoga
38	613.5	26.1	459	16	Q8ZCQ1	Q8ZCQ1 salmonella
39	606.5	25.8	452	16	P73354	P73354 synechocyst
40	606.5	25.8	481	16	Q9LEK0	Q9LEK0 shigella so
41	604.5	25.7	491	2	Q9PGL3	Q9PGL3 xylella fas
42	603.5	25.6	371	2	Q31388	Q31388 bradyrhizob
43	600.5	25.5	416	16	P73940	P73940 synechocyst
44	600	25.5	460	2	Q06439	Q06439 rhodobacter
45	589	25.0	389	16	Q9HAX1	Q9HAX1 pseudomonas

ALIGNMENTS

RESULT 1
ID Q9JVT1 PRELIMINARY; PRT; 499 AA.
AC Q9JVT1
DT 01-OCT-2000 (TREMUREL.15, Last sequence update)
DT 01-OCT-2000 (TREMUREL.15, Last sequence update)
DT 01-JUN-2002 (TREMUREL.21, Last annotation update)
DE Putative periplasmic serine protease (EC 3.4.21.).
GN NMA0710.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761915;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S., Jagers K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."
RT Nature 404:502-506(2000).
RL -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC EMBL: AL162754; CAB83996.1; -.
DR EMBL: AL162754; CAB83996.1; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00595; PDZ; 2.
DR PRINTS: PRO0834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PSS0106; PDZ; 1.
DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
KW Hydrolyase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 499 AA; 52563 MW; 73E979F0A7F5B969 CRC64;

Query Match	Similarity	95.5%	Score	2246	DB	16	Length	499
Best Local	Similarity	95.9%	Pred. No. 3	3e-129				
Matches	444	Conservative	6	Mismatches	13	Indels	0	Gaps
Qy	1	VPKKYQYFALAAALCAALLAGCEKASFFGADCKEASFYERIEHTYKDDGVSMLDPFAOL	60					
Db	1	MFKKQYVLAALCAALSLAGCDKASFFGADCKEASFYERIEHTYKDDGVSMLDPFAOL	60					
Qy	61	VQSEGPVAVNIQAAPAPRTONGSSNAETDSDPLADSDPFYEFKRLVPMPEIPOEADD	120					
Db	61	VQSEGPVAVNIQAAPAPRTONGSSNAETDSDPLADSDPFYEFKRLVPMPEIPOEADD	120					
Qy	121	GLNPGSGFIIISKNGYILTNTHVVGMSIKYLLNDKREYTKLIGSDVQSVALLKTKIDA	180					
Db	121	GLNPGSGFIIISKNGYILTNTHVVGMSIKYLLNDKREYTKLIGSDVQSVALLKTKIDA	180					
Qy	181	TEELPVVXIGNPKMLKPEEWAAIAPGPFDSVTAGIVASAKRSLPNESTYPIQTDDVA	240					
Db	181	TEELPVVXIGNPKMLKPEEWAAIAPGPFDSVTAGIVASAKRSLPNESTYPIQTDDVA	240					
Qy	241	INPGNSGGLPFLNLKQGVVINSQIYSRSGGFMSIPAIPIDVAVNAEQLKNTGKYVRQO	300					
Db	241	INPGNSGGLPFLNLKQGVVINSQIYSRSGGFMSIPAIPIDVAVNAEQLKNTGKYVRQO	300					
Qy	301	LGVIITOEVSYGLAOSFGHDKASGALIAKILPSSPAPRGLAQGDIVLSDGEIRSSGDL	360					
Db	301	LGVIITOEVSYGLAOSFGHDKASGALIAKILPSSPAPRGLAQGDIVLSDGEIRSSGDL	360					
Qy	361	PVMYGAITPGKEVSLGVWRKBEITIKAKLGAAEHTGASKTDEAPYEQSGTFVSYS	420					
Db	361	PVMYGAITPGKEVSLGVWRKBEITIKAKLGAAEHTGASKTDEAPYEQSGTFVSYS	420					
Qy	421	AGITLQTHTDSSGKRLVVYRVSDAERAPGLRHDEIILAVRASP	463					
Db	421	AGITLQTHTDSSGKRLVVYRVSDAERAPGLRHDEIILAVRASP	463					
RESULT 2								
Q9ALSL1	PRELIMINARY	PRT	474	AA				
AC	Q9ALSL1							
DT	01-JUN-2001 (TremBLrel. 17, Created)							
DT	01-JUN-2001 (TremBLrel. 17, Last sequence update)							
DT	01-MAR-2002 (TremBLrel. 20, Last annotation update)							
DE	MUCD.							
GN	MUCD.							
	Pseudomonas aeruginosa.							
	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;							
	Pseudomonas.							
OX	NCBI_TaxID=287;							
RN	(1)							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=UCBPP-P14;							
RA	Yorgey P.S., Rahme L.G., Tan M., Anubel F.M.,							
RT	"The Roles of mucD and Alginate in the Virulence of Pseudomonas							
RL	aeruginosa in Plants, Nematodes, and Mice.";							
DR	Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.							
DR	EMBL; AF343973; AAK11276.1; -							
DR	InterPro; IPR001478; PDZ.							
DR	InterPro; IPR001940; ProteaseC.							
DR	InterPro; IPR001254; Ser_protease_Try.							
DR	Pfam; PF00595; PDZ; 2.							
DR	Pfam; PF00089; trypsin; 1.							
DR	PRINTS; PR00834; PROTEASES2C.							
DR	SMART; SM00228; PDZ; 2.							
DR	PROSITE; PSS0106; PDZ; 2.							
DR	PROSITE; PSS0240; TRYPsin_DOM; 1.							
KW	Hydrolase; Serine protease.							
SO	SEQUENCE	474	AA	A22FD4338B859D4C	CRC64			
Query Match								
Best Local	Similarity	52.1%	Score					

QY	5	LEPFAQVQSEGPAAVNI---QAAPAPRTONGSGNAETSDPLADSDP-FYE.	110	MEPIE---QEADDDGGLNFGSGFTISKNGYILTNTNHYVAGMGSIKYLNDKREYI	167	SDVQSDVALLKIDATEELPVVKIKGNKONIKPGWVAIAAPGPFQDSVYAGVSAKGRSL	222	ADPSSDVAIVKITA-KULPTLKLGDNSKCLKVGEWVAISPGFQDSVYAGVSAKGRSL	287	AEOLKTKGVQKQOLGVIIIOEVSYGLAQSPGLDKASGALLIAKLPGSPAPRAGLOAGDIV	347	LSLDGGIRSRSGPLPMVMAITPGKEVSLGWKKEEITIKKLGAAHHTGASSKTD-	406	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	466	LSLNGSGINESALPHLVGMMKRGDINIDVIRNGRKSLSMAVGLPV-----DDEE	528	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	588	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	648	LSLNGSGINESALPHLVGMMKRGDINIDVIRNGRKSLSMAVGLPV-----DDEE	708	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	768	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	828	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	888	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	948	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1008	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1068	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1128	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1188	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1248	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1308	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1368	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1428	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1488	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1548	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1608	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1668	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1728	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1788	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1848	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	1908	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	1968	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2028	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2088	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2148	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2208	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2268	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2328	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2388	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2448	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2508	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2568	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2628	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2688	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2748	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2808	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2868	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	2928	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	2988	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3048	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3108	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3168	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3228	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3288	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3348	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3408	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3468	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3528	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3588	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3648	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3708	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3768	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3828	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	3888	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	3948	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4008	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4068	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4128	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4188	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4248	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4308	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4368	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4428	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4488	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4548	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4608	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4668	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4728	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4788	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4848	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	4908	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	4968	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	5028	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	5088	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	5148	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	5208	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	5268	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	5328	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	5388	ADOLKTKGKVSRLMGVIOEVKDLAESFGDLKPSGALVAOLVEDGPAKAGGLGVGDVI	5448	-----APYTEQSGFTSPVSAGITLQTH--TDSGSHLVVVRVSDA-AERAGLRHGD	
----	---	---	-----	---	-----	--	-----	---	-----	--	-----	--	-----	---	-----	--	-----	---	-----	---	-----	--	-----	---	-----	---	-----	---	-----	---	-----	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	--

RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Seiler M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RA "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 RT opportunistic pathogen,"
 RT Nature 406:959-964(2000).
 RU Nature 406:959-964(2000).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: U949151; AAC43718.1; -
 DR EMBL: U32853; AAC43676.1; -
 DR EMBL: AB004511; AAC04155.1; -
 DR InterPro: IPR001478; PDZ; 2.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR HydroLase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 474 AA; 50321 MW; 8AB5D2A8967BEBE CRC64;

Query Match 41.9%; Score 985.5; DB 16; Length 474;
 Best Local Similarity 52.5%; Pred. No. 2.7e-52;

Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

54 LPDPAQLVQSEGPAYVNI---QAAPARTQNGSNAETSDPLADSDP-FYEFKFLVNI 109
 27 LPDTPPLVQASPAVNIISTROKLP---DRAMAGQLSPLEGLPMPFRDPLERSIQ 82
 110 MPEIP---QEADDOGLNFGSGFIISKGYIITNTHVVAQMSIKVLYNDKREYTKLIG 166
 83 VPRNPRGQRAQ---SLGSGFIISNDGYILTNHNVADADEILVRLSDRSHKAKLIG 138
 167 SDVQSDVALLKIDATEELPVYKIGPKNLKPGEWAAIGAPFPNSVTAGIYSAGRSI 226
 139 ADPSDAVAVLIEA-KNLPFLKGDGSMKLKGEWVLAISPGFPHSVTAIVSAGRSI 197
 227 PNESTYPTIOTDVAINPGNSGGLFNLKGQVVGINSQIYSRSGFMGIFPAIPIDVAMV 286
 198 PNESTYPTIOTDVAINPGNSGGLFNLKGQVVGINSQIYSRSGFMGIFPAIPIDVAMV 257
 287 AEOLKNTGKVGQGLVLIQEVSYGLAOSFGLDKASGALIAKILPGSPAPERAGLQAGDIV 346
 258 ADQKKAGKVGKGLVLIQEVSYGLAOSFGLDKASGALVADLVEDGPAKGLQYGDVI 317
 347 LSLDGGERTSSGDLPMVVGATTPGKEVSLGWRKGEITTIYAKKLGNAAEHTGASSTDE 405
 318 LSLNGOSINESADLPHLVGNMKPGDKINIDVTRNGQRKSLMAVAGSLP-----DDEE 370
 406 -----APTEGOSGTFVESAGITLQTH--TDSGKHLVVAVSDA-ABRAGLRHGEI 456
 371 IASMGAPAEKSSNRGLVTVADLTABQRKSLDIQG-GVVIKEVQDPPAVITGLRPGDVI 428

RESULT 4

08XPT5 PRELIMINARY; PRT; 490 AA.
 AC 08XPT5 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Probable protease signal peptide protein (EC 3.4.-.-).
 GN RSP1552 OR RS02108.
 OS *Ralstonia solanacearum* (*Pseudomonas solanacearum*).
 OC *Ralstonia solanacearum*.
 CC Bacteria; Proteobacteria; beta subdivision; *Ralstonia* group;
 OC *Ralstonia*.

OX NCBI_TaxID=305;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=GM1000.
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangelot S.,
 RA Aplat M., Billault A., Brottier P., Camus J.C., Catalicio L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.,
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*,"
 RT Nature 415:497-502(2002).
 DR EMBL: AL646085; CAD18703.1; -
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR HydroLase; Plaemid; Complete proteome.
 SQ SEQUENCE 490 AA; 50606 MW; AD54255965AA4888 CRC64;

Query Match 38.2%; Score 900; DB 16; Length 490;
 Best Local Similarity 47.6%; Pred. No. 4.6e-47;
 Matches 195; Conservative 65; Mismatches 134; Indels 16; Gaps 7;

56 DPAQLVQSEGPAYVNIQ-AAAPARTQNGSNAETSDPLADSDPFEYFELVPMPEIP 114
 59 DPAIYERYPAYVNIISTTARAQRT---SIQGLPVGVSDDPFAEFRRP---MPQAP 110
 115 QEADDOGLNFGSGFIISKGYIITNTHVVAQMSIKVLYNDKREYTKLIGSDVQSYA 174
 111 QQQGDQVVKLGSGFIVSPGLILTNHNVADQEVSVKLTDRREFKAKVIGVQKQSDVA 170
 175 LKIDATEELPVYKIGPKNLKPGEWAAIGAPFPNSVTAGIYSAGRSIIPNESYPT 234
 171 VLRI-AAANLPYQIGSPAKTKGEPVLAISPGFENTYTAGIYSAGRSIIPDDTYVP 229
 235 IQTDVAINPGNSGGLFNLKGQVVGINSQIYSRSGFMGIFPAIPIDVAMVABQLKNTG 294
 230 IQTDVAINPGNSGGLFNLKGQVVGINSQIYSRSGFMGIFPAIPIDVAMVABQLKNTG 289
 295 KYQKQGLVLIQEVSYGLAOSFGLDKASGALIAKILPGSPAPERAGLQAGDIVSLDGEI 354
 290 KYRGRGLGISVQGVDSGLADSFNLKPEGALVNAVERKDPFAKAGLQPGDIVLIQIGDVI 349
 355 RSSGDLPMVVGATTPGKEVSLGWRKGEITTIYAKKLGNAAEHTGASSTDEAPTEGQSG 414
 350 GHSGLDPEQVAELIKPGSTVPLQITIRKGPYALSVTVBAND-AKYAANTSAP---DKG 404
 415 TFEVESAGITLQTHDTS-GKHLVVAVSDAERAGLRHGEIILAVRASP 463
 405 RLGLAVRPLQAEKRGSGLPGLVVMSSGPAKAGLQPGDIVLISINTP 454

RESULT 5

08Y016 PRELIMINARY; PRT; 505 AA.
 AC 08Y016 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Probable periplasmic protease signal peptide protein (EC 3.4.-.-).
 GN RSC1058 OR RS04156.
 OS *Ralstonia solanacearum* (*Pseudomonas solanacearum*).
 OC *Ralstonia solanacearum*.
 CC Bacteria; Proteobacteria; beta subdivision; *Ralstonia* group;
 OX NCBI_TaxID=305;
 [1]

RN [1]
 RP SEQUENCE FROM N.A.
 RA Heu C.-C., Shieh S.-Y., Yang M.-T.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 SO EMBL: AF46385; AAL74147.2; -- 368PB263FCB0F713 CRC64;
 SEQUENCE 511 AA; 52284 MW; 368PB263FCB0F713 CRC64;
 Query Match 37.6%; Score 884; DB 2; Length 511;
 Best Local Similarity 47.4%; Pred. No. 4.6e-46;
 Matches 203; Conservative 59; Mismatches 118; Indels 48; Gaps 12;
 QY 54 LPDPAQLVQSGPAAVNIQA-----PADRTONGSGNAETSDPLADSPFFYFFRL-- 106
 DB 38 LPDFTNLVEQVGPVNIETTTTRKDAMARSQ--GPGRGCGGAMPDDQMPFFFRPRG 96
 QY 107 -----VPNMP-EIPOSEADGCL--NFGSGFIISKNGYLLTTHVAVAGMSITVLLNDR 158
 DB 97 PDFQMPGPGPGGGDDGGIAGKSGSGFTISADGYVLTNNHVVDGASEVTVKLTDRR 156
 QY 159 EYTAKLISDVQSDVALLKIDATEELPVYKIGNPKULKPEGVVAIAGAPGFNDSTYAGI 218
 DB 157 EFKAKVQSDQDFVALKIEA-KGLPTVRIGDSNTLKFGQNVVAIGSPFGLDHSTAGI 215
 QY 219 VSAGKRSIP--NESYTPPIQTDVAINPNSGAPLFNLKQVVGINSQIYSRSGFGNIGSF 276
 DB 216 VSATGRSNPVADQRYVPFIQTDVAINQNSGGPLNTRGEGVINGISQIFSAGYWGISF 275
 QY 277 AIPIDVANNVAEQLKNTGKVGQGLVITIOEVSYGLAQSGFGLDKASGALIAKLIPSPAP 336
 DB 276 AIPIDLASAEQIKASGVNSGMLGVAAGVPIDLKAQGLGIPDTRGALVNDIPAGSPAG 335
 QY 337 RAGIQAQGVIVSLDGEIRSGDLPMVVGATTPGKEVSLGVWRKGEIITKALGNAEH 396
 DB 336 KAGIEVGVIVSVNGKEIAVASDLEPMITGMPGTIVSLNLRDGRQVYTVLGLIENE 395
 QY 397 TGAASKTDEAPY---EQSGSTFVSAGITLQTHTDSGKGLVYRVSD--AAERA-- 448
 DB 396 SGSS-----APRTAADSKSPASAVELLGL-----QVADLTAERSRN 434
 QY 449 GLRRGDEI 456
 DB 435 GLEAGEGV 442
 RESULT 8
 Q9PBA3 PRELIMINARY; PRT; 514 AA.
 AC Q9PBA3;
 P 01-OCT-2000 (TREMBlrel. 15, Created)
 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Periplasmic protease.
 GN Xf2241.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365117; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Apentio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.B., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove U.M., Britones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carreir H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto B., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Pereira V.C.A., Perro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohensei J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kurame E.B., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos B.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nacimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Paixoto R.B., Pereira G.A.G., Pereira H.A. Jr., Pasquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
 RA da Silva A.C.R., de Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siguelia M.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Teshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 CC -
 DB EMBL: AE004037; AAF85040.1; --
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser. protease Try.
 DR InterPro: IPR00126; Ser. protease V8.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; tryptsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR PRINTS: PR00839; V8PROTEASE.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 DR Hydrolase; Serine protease; Complete proteome.
 SO SEQUENCE 514 AA; 54140 MW; 707C23FD3C82BAC CRC64;
 Query Match 36.7%; Score 862.5; DB 16; Length 514;
 Best Local Similarity 46.2%; Pred. No. 9.6e-45;
 Matches 194; Conservative 64; Mismatches 129; Indels 33; Gaps 10;
 QY 54 LPDPAQLVQSGPAAVNIQAAPRTONGSGNAETSDPLADSPFFYFFRL-----V 107
 DB 54 LPDFTNLVEQVGPVNIETTTTRKVGKRGIFLND-----IPEFFRFFPGDPDM 106
 QY 108 PNMPEIPOSEADGCL--NFGSGFIISKNGYLLTTHVAVAGMSIKVLLNDRREYTKL 164
 DB 107 PNOPRGQD--DEGIGRAGWGSGFIISKNGYLLTNNHVIVGASBVTIKLTDREFFAKI 164
 QY 165 IGSDVQSDVALLKIDATEELPVYKIGNPKULKPEGVVAIAGAPGFNDSTYAGISAGR 224
 DB 165 IGSDQVDVALLKIDA-KMLPTVRIGDSSTLSKSGQVVAIGSPFGLDHSTYAGISALGR 223
 QY 225 SLPNES-YTPIQTDVAINPNSGAPLFNLKQVVGINSQIYSRSGFGMGISFAIPIDVA 283
 DB 224 STSDQRYVFIQTDVINGNSGAPLFNLTRGVYIGINSQIFSASGGMGISFAIPIDVA 283
 QY 284 MNVAEQKNTGYQVQGLGVITIOEVSYGLAQSGFGLDKASGALIAKLIPGSPAERAGIAG 343
 DB 284 INNAEQIKRTGKQVQSMIGVEIGPIDLKAQGLPLPSRGALVNNIIPHSAPAKAGIEVG 343
 QY 344 DTVLSLDGGEIRSGDLPMVVGATTPGKEVSLGVWRKGE--EITIKALGNAEHTGASS 401
 DB 344 DVIRSVNGKVISSFSDDLPLIGMPPGTATLGIIRKREIIVTSLALN-QDATTASBD 402
 QY 402 KTBAPYTBEOGSGTFSVSAGITLQTHTDSGKGLVYRVVAVDAERAGLRHGDDELAVRA 461
 DB 403 EDTTAPKERTSA--NVELLGLQVENLSAARERL-----ASSQAKAGVRLTAVTA 452

RESULT 9
 Q98CS8 PRELIMINARY; PRT; 513 AA.
 AC Q98CS8;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Serine protease, HtrA/degQ/degS family.

OC Bacteria, Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 RN NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Secubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woe L.,
 RA Chen Y., Paulsen I.T., Bisen J.A., Karp P.D., Boyee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gallet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Semphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Rao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neeter E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 CS8.";
 RL Science 294:2317-2323 (2001).
 RL [2]
 RL SEQUENCE FROM N.A.
 MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlet K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Planagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens CS8.";
 RL Science 294:2323-2328 (2001).
 RL EMBL; AB009062; AAL4191.1; -
 DR EMBL; AB008028; AAK6786.1; -
 KM Protease; Complete proteome.
 SQ SEQUENCE 523 AA; 54436 MW; AFA52B8A5E258ED CRC64;
 Query Match 31.2%; Score 735; DB 16; Length 523;
 Best Local Similarity 37.0%; Pred. No. 66-37;
 Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;
 QY 36 SFVERIETKDDGSVSMILPPFAQLVOSBGPVAVNIQAAPARTONGSNAETDS----- 90
 DB 44 SFAAPVETABQ-----VPSFANVADVAVSPAVSV-----RVGSNVPASDSSNFSF 91
 QY 91 -----DPLADSDPFYEFKELVNMPEIPOEABDDGLN-----FGSGPI 130
 DB 92 NFGGRGLDQLPDHDKRFFKEFGQ-----NQDSRGPNNHRODKGPLRPAQSGSF 146
 QY 131 ISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSDVQSDVALIKIDATELPPVKIG 190
 DB 147 ISEBGVVTNNHVVDDGSAVTVVMDGTELEAKLVGRDPRDLDLALKVNVNKKFTYVKA 206
 QY 191 NPKULKGEVVAALGAPFGFNSVTAGIVSAKGRSLPNESYTPFIQTDAVNPNGSGPL 250
 DB 207 DDTIRIVDDWVAAGVPGFGIGTIVSGIISARGRDIGSGPYDYLQIDAANRNSGGBA 266
 QY 251 FNLGGOVVGINSQIYSRSGFEMGISFAIPIDVAMVABQLKNTGKQVQRLGVIIQVSV 310
 DB 267 FNLGGEVVGINTALFSPSGGVNGIAPALPSSVAKVIALDLQDQKVERKMLGVLQIPVSK 326
 QY 311 GLAOSFGIDKASGALIAKILPGSPAERAGLQAGDIVLSLDSGEIRSSGDLPMVVGAIITGP 370
 DB 327 DIASLIGLAEKAGLVSPQSGSPGDKGIGKIGDIIITVANGPVPVADARLSRIGMAIN 386
 QY 371 KEVSLGVWRKEBEITIKALGN--AAHTGASKTDEADYEQSGTFSVASGITLQTH 428
 DB 387 SKVVISLWRGKSGSVTTLGDLTDEASKATPSONDDKSGSGSSEKVLSSLGLTVSPS 446
 QY 429 TDSGKHLVVVRV---SDAABRAGLRHGDILLAV 459
 DB 447 DDGNG--LAITVDVDDSDAAAR-GLTKGTITSV 477

RESULT 12
 Q985F9 PRELIMINARY; PRT; 516 AA.
 ID Q985F9;
 AC Q985F9;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Serine protease.
 GN MR7692.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria, Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Mesorhizobium.
 RN NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasaemoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AP003012; BAB54103.1; -
 DR MEROPS; S01.273; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR000126; Ser_proteas_V8.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASE2C.
 DR PRINTS; PRO0839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 KM Protease; Complete proteome.
 SQ SEQUENCE 516 AA; 53704 MW; FE8DAADC099ABDC6 CRC64;
 Query Match 30.6%; Score 719.5; DB 16; Length 516;
 Best Local Similarity 36.2%; Pred. No. 5.2e-36;
 Matches 174; Conservative 80; Mismatches 177; Indels 49; Gaps 9;
 QY 10 LALCALALAGCKKAGSFADKKEASFVERIETKDDGSVSMILPPFAQLVOSBGPV 69
 DB 16 LAAAVAVAAVAVIGVAGALISGTSPLADAVRVEAVQVQ-----FADVERVSPAV 67
 QY 70 N--IQAAPARTONGSNAETDSDPLADSDPFYEFKELVNMPEIPOEABDDGLN--- 124
 DB 68 SVAVKAKIQPTADGSDQ--QDGRDNLPNPQLRRFFKEF-----RGFGDGGQNDG 118
 QY 125 -----FGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIG 166
 DB 119 HRRFGRHRSNDQPRPAVQSGFFISBDGYLVNTHVVEEGTALVTJNDKEIDAKLVG 178
 QY 167 SDVQSDVALIKIDATELPPVKIGNPKULKGEVVAALGAPFGFNSVTAGIVSAKGRSL 226
 DB 179 TDRITLAVLAKVGGGFTYVDPADDSKVRGVAMVAVGNPFGIGGVTAIVSARGDI 238
 QY 227 PNESTYTPFIQTDAVNPNGSGPLFNLKGOVVGINSQIYSRSGFEMGISFAIPIDVAMV 286
 DB 239 GAGPYDPLQIDASVNRKNSGGFTFNLNGVVGINTALFSPSGSVGIAPIPASTACV 298
 QY 287 AEQLKNTGKQVQRLGVIIQVSVGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346
 DB 299 VEDLMKSGAVORGMVLGVEIOPVTSDIASISGLKSNNGALVSSAADDPGKKAGITAGVI 358
 QY 347 LSLDSGEIRSSGDLPMVVGAIITPKEVSLGVWRKEBEITIKALGN--AAHTGASKTDE 405
 DB 359 TVEGKDVAVSPKELALIGAVSPKSVDTVWBDGKSQITIKVDLKGKLPASDKQASNDQ 418
 QY 406 APYTEQSGTFSVASGITLQTHTDSGKHLVVVRV---SDAABRAGLRHGDILLAVRAS 462

Db	419	QPAAPAKPDTL----	ADLGLRVTSKSENGKGLVVTDDVDEPSAAADR-GIOPGIIITAVNSN	473
RESULT	13			
ID	08UDS7	PRELIMINARY;	PRT; 514 AA.	
AC	08UDS7;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Serine protease DO-like			
GN	HTPA OR ATU2043 OR AGR_3700.			
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Rhizobiaceae; Rhizobium.			
OX	NCBI_TaxId=176299;			
RN	[1]			
	SEQUENCE FROM N.A.			
	MEDLINE=21608550; PubMed=11743193;			
	Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,			
	Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,			
	Chen Y., Paulsen I.T., Eilen J.A., Karp P.D., Boye D. Sr.,			
	Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,			
	Kutyavin T., Levey R., Li W.-J., McCilland E., Palmeri A.,			
	Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,			
	Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepan W., Perry M.,			
	Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,			
	Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,			
	Neeter E.W.;			
RT	"the genome of the natural genetic engineer Agrobacterium tumefaciens			
RT	C58."			
RL	Science 294:2317-2323 (2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21608551; PubMed=11743194;			
RA	Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,			
RA	Qurollo B., Goldman B.S., Cao Y., Aekenzai M., Halling C., Mullin L.,			
RA	Houmel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,			
RA	Mollam C., Allinger M., Doughly D., Scott C., Lappas C., Markelz B.,			
RA	Planegan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,			
RA	Cielo C., Slater S.;			
RT	"Genome sequence of the plant pathogen and biotechnology agent			
RT	Agrobacterium tumefaciens C58."			
RL	Science 294:2323-2328(2001).			
DR	EMBL; AB009155; AAL3035.1; -			
	EMBL; AB008120; AAK8796.1; ALT_INIT.			
	Protease; Complete proteome.			
	SEQUENCE 514 AA; 53551 MW; 1138B1328459EA7 CRC64;			
Query Match	30.4%;	Score 716;	DB 16;	Length 514;
Best Local Similarity	36.4%;	Fred. NO. 8.4e-36;		
Matches 170;	Conservative 75;	Mismatches 146;	Indels 76;	Gaps 10;
QY	48	GSVSMILPDPFQALVQSEGP-----AVNVIQAAPRTONGSGNAETDSDPLAD	95	
DB	24	GSLS---APVAARQSHQSPESVADLAELDLAVVANNISQVKNIE-GKGPV---PKLPE	76	
QY	96	SDPFYEFFKRLVPMNPETIQSEADQG--LNFSGGLTISKNGYILNTTHVAVKSGSIKV	152	
DB	77	GSFPQEEFFK-----DYFDSQKEPGEKVNLSGSGFYIDAGYVVNNHYIEGADAIIEV	129	
QY	153	LMDKRETTAKLTIGSDVQSDVALLKIDATEELPLPVKGNPKIKRGVAVNAGAPFGDN	212	
DB	130	IFPPGSKLKATLVGTGDTKTDLSVLKVEKEKTPKAKKPFSDSSMRIGDVMVAVGNPFGIGG	189	
QY	213	SVTAGIYSAKGRSLPNESYTPFIQTDVAINFGNSGGLPFNLKQGVVGNISQIYSKSGGFM	272	
DB	190	SLTGVGISARCRNINAGPYDNFIQTDAIINKNGSGGLPFNNKGEVIGINTAIISPSGSI	249	
QY	273	GISPAIPIDVANMVAEQLKATGKVRQROGLYITIEGVSAIGLAQSPFLDKASGALLAKILPG	332	
DB	250	GIGAVAPTELAQNTVOOLIEFGETRFRGWLGVAVOPVPTDDVAASIGMDSAKALLISGVAKG	309	

QY	333	SPERRAGLQAGDITLSDGGEIRSSGGLPWWVGAITPGEKVSILGWKKGEITTKALGN	392
Db	310	GPVENGTQAGDVAIVLAKPKDGDKQDINEMRDLRIVAESPVGKEVDVVYDDGKEETVAKYLQ	369
QY	393	AAEHTGASKTDEAPYTEQOQG-----TFSESAGITLQTHDSSGK	434
Db	370	LQDTTBKASRTD-PQGEDGGGSMVAPDDKDGDDQADQDPPEVREAPQTV-----LGM	422
QY	435	HLVY-----VRSDAARAGLRHDELLAV	459
Db	423	NLVVLSNELRTEKGIASVEGVLVASVDPSGPAQKQMKAGDIVLEV	469
RESULT 14			
08YG32	08YG32	PRELIMINARY;	PRT; 513 AA.
AC	08YG32		
DT	01-MAR-2002	(Tremblrel, 20, Created)	
DT	01-MAR-2002	(Tremblrel, 20, Last sequence update)	
DT	01-JUN-2002	(Tremblrel, 21, Last annotation update)	
DE	Protease DO (EC 3.4.21.-)		
GN	BMEI130.		
OS	Brucella melitensis.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC	Brucellaceae; Brucella.		
OX	NCBI_TaxId=28459;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=16M / ATCC 23456 / BIOTYPE 1;		
RX	MEDLINE=20020109; PubMed=1175668;		
RA	DelVecchio V.G., Kaparatel V., Redkar R.J., Patra G., Mujar C., Los T.,		
RA	Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,		
RA	Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,		
RA	Seikov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-O.,		
RA	Haelekom R., Kyripides N., Overbeek R.;		
RT	"The genome sequence of the facultative intracellular pathogen		
RT	Brucella melitensis.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).		
DR	EMBL; AE009571; AA152511.1; -.		
DR	InterPro; IPR001478; PDZ.		
DR	InterPro; IPR001940; Protease2C.		
DR	InterPro; IPR001254; Ser_protease_Try.		
DR	Pfam; PF00595; PDZ; 2.		
DR	Pfam; PF00089; trypsin; 1.		
DR	PRINTS; PR00834; PROTEASES2C.		
DR	SMART; SMO0228; PDZ; 2.		
DR	PROSITE; PS50106; PDZ; 2.		
KM	Hydrolase; Complete proteome.		
SR	SEQUENCE 513 AA; 53514 MW; C472PEF99DFC6268 CRC64;		
Query Match 29.5%; Score 693.5; DB 16; Length 513;			
Best Local Similarity 37.4%; Pred. No. 2e-34;			
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;			
QY	9	ALAAALCAALLAGECKAGSPFGADKKEASFYERTENTYDDGGSVMILPD---FAOLVQSEG	65
Db	14	AAVAASAA-----AGAFVVTGRLGLALNERRA-----AVHTTPQAGFALVLEKVR	61
QY	66	PAVVNIQAAPRPTONGSGN---AETSDPLADSDFPEYFF-----KRLV	107
Db	62	PAVSVSVKKDQVQETSNRGPQFGPQFPQDLRDPGHLPRKFRPFGMEPRGDSRDNRRGK	121
QY	108	PNMEITQGEABDQGLNFGSGFIISKNGYLITNHHVVAWGMSITVLNDRKREYAKLIGS	167
Db	122	ANKPRPGEHER---VAQSGGFVISEGGYVAVTNHHVVSDDGDAYTVVLDGTELDKALIGA	177
QY	168	DVQSDVALILKIDATE-ELPVPVKIGNPKNLKGENVAAIGA-PFGFDSNVTAGIVSAKGRSL	226
Db	178	DPRDIDLAVLKNAPKRKFVVAEFDMDKXAVGVGVAVGNPFGIGRGIVTSISVABRGDI	237
QY	227	PNESTYPIOTDVAIVPQNSGGPLFNILKGOVVGINSOIYSRSGGFMGISFAIPIDIVAMV	286

Db 238 GAGPYDDIQTIDAANKSGGAPDLSGEVIGINTAIFSPSGSVGIAFAPISSTAKOV 297
QY 287 AEOLKNTKGVQGLGVIIQEVSYGLAOSFGLDKASGALIAKILPSPAPRAGLOAGDIY 346
Db 298 VDQLTKKGSVERGWIQVQPTKIDIAISIGLAEEKGAIYASPODDGPAKAKIKKGDVI 357
QY 347 LSLDGEIRSSGDLVVMVCAITPGKEVSLGVWR--KGEETITK-ALGNAAEITGASSKT 403
Db 358 TAVNGEIVQDPEDLARKVANIAFGEKVALITWRKNAEEINVTIAMPMFDKSGSGQSDND 417
QY 404 DEAPYTEOQSGTFVESAGITLTQHTDSSGKHLVYRV--SDAARAGLRHGDIELAV 459
Db 418 NDGGGE-----TLDSTGLTVVPSRD--GKGIVVTVDPDSDAADR-GIRSGDIIVSV 467

RESULT 15

08YI32 PRELIMINARY; PRT; 524 AA.
AC 08YI32.
DT 01-MAR-2002 (TRENBLrel. 20, Created)
01-MAR-2002 (TRENBLrel. 20, Last sequence update)
01-JUN-2002 (TRENBLrel. 21, Last annotation update)
Protease DO (EC 3.4.21.-).
BME10613.
GN Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
CX NCBI_TaxID=29459;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapatali V., Redkar R.J., Patra G., Mujer C., Los T.,
R Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
R Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman B.,
R Selkov E., Elizer P.H., Hagius S., O'Callaghan D., Leeson J.-J.,
R Haelekom R., Kyripides N., Overbeek R.;
RA "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009503; AAL51794.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR HydroLase; Complete proteome.
SEQUENCE 524 AA; 55273 MW; 15B98E245223658 CRC64;

Query Match 29.0%; Score 683; DB 16; Length 524;
Best Local Similarity 40.5%; Fred. No. 8.9e-34;
Matches 161; Conservative 67; Mismatches 152; Indels 18; Gaps 7;

QY 48 GSVSMILP-----DPAQLVQSEGPVAVNIQAAPARTONGSGNAETSDPLADSDP 98
Db 26 GTVMGTPTPALAAGPASYADLAEGLDVAVNISTQYVK-DDGSDGVPMPQVEGSP 84
QY 99 FYEFPKRLVPMNPRLPOEBADGDLNFGSGFTI-SKNGYILTNTVAVAGWSIKVLLNDK 157
Db 85 FOEFFKDFP-NDKDAQGDGDSRKVQSLSGFTIIDAEKGYITNNNHVIAADADEIVNFDG 143
QY 158 REYTAKLIGSDVQVALIKITDTE-ELFVYKIGNPKNLKPGEWVAIGAPRGFDNSVTA 216
Db 144 SKLAELVQKOTKTLAILKVDPSKRKLAVHFGNSEKARIGDMLAIGNPFGLGIVTA 203
QY 217 GIVSAKSLPMBESTYPTFIQTVAINPNSGGPLFVLKGVQVGINQIYRSRSGGFMGISF 276
Db 204 GIISARKDINSGRPYDDFQITPAINRNGSGPLFMDGKVGIGINTAITSPPSGSIGIGF 263
QY 277 AIPIDVANNVAEQLKNTKGVQGLGVIIQEVSYGLAOSFGLDKASGALIAKILPSPAP 316

Db 264 ALPAMWAGVITDQKERGEVRRGWLGVRLQPTVEDIAQSLGKTKGALIGLIENSGVD 323
QY 337 RAGLOAGDIVLSLDGEIRSSGDLPMVWVGAITPGKEVSLGVWRKGEETITKALGNAAEH 396
Db 324 NKALBAGDVVIRPDGKRVDTARDLPLVAERPVGKEVEIVVIRQAEKTLKKGCLVPE- 382
QY 397 TGASSKTDEAPYTEOQSGTFVESAGITLTQHTDSSGK 434
Db 383 ---DDKSTPEAVEDQVPAPDDGEGPGARQET-PDKSDK 416

Search completed: July 11, 2003, 10:33:11
Job time : 87 secs

